This Page Is Inserted by IFW Operations and is not a part of the Official Record

BEST AVAILABLE IMAGES

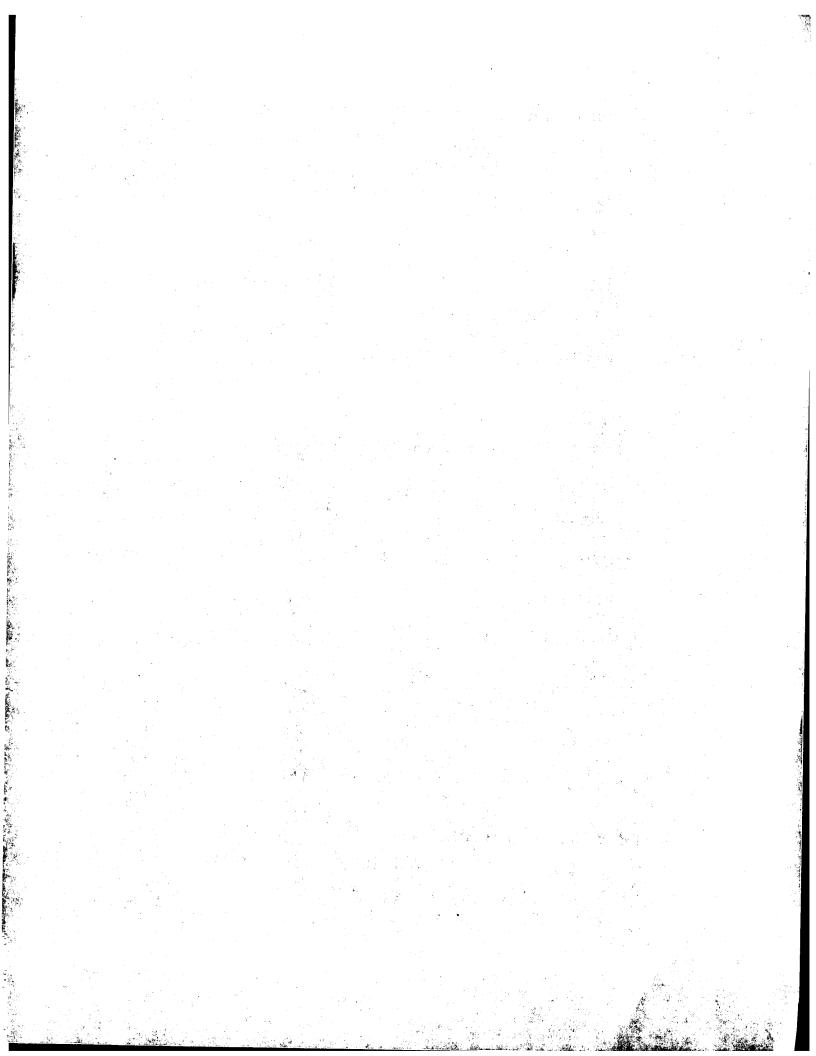
Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents will not correct images, please do not report the images to the Image Problem Mailbox.



PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classificati n ⁶:

70,

(11) International Publication Number:

WO 99/04008

C12N 15/40, C07K 14/18, C12Q 1/70, C07K 16/10, A61K 39/29

A2 (43) Int

(43) Internati nal Publication Date:

28 January 1999 (28.01.99)

(21) International Application Number:

PCT/US98/14688

(22) International Filing Date:

16 July 1998 (16.07.98)

(30) Priority Data:

60/053,062

18 July 1997 (18.07.97) US 27 January 1998 (27.01.98) US

09/014,416 27 January 1998 (27.01.98) US

(71) Applicant: THE GOVERNMENT OF THE UNITED STATES OF AMERICA, asrepresented by THE SECRETARY, DEPARTMENT OF HEALTHAND HUMAN SERVICES [US/US]; Office of Technology Transfer, National Institutes of Health, Suite 325, 6011 Executive Boulevard, Rockville, MD 20852 (US).

(72) Inventors: YANAGI, Masayuki; 257 Congressional Lane, #402, Rockville, MD 20852 (US). BUKH, Jens; 7 Center Drive, MSC 0740, Bethesda, MD 20892 (US). EMERSON, Suzanne, U.; 18201 Woodcrest Drive, Rockville, MD 20852 (US). PURCELL, Robert, H.; 17517 White Grounds Road, Boyds, MD 20841 (US).

(74) Agents: FEILER, William, S. et al.; Morgan & Finnegan, L.L.P., 345 Park Avenue, New York, NY 10154 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND USES THEREOF

(57) Abstract

The present invention discloses nucleic acid sequences which encode infectious hepatitis C viruses and the use of these sequences, and polypeptides encoded by all or part of these sequences, in the development of vaccines and diagnostics for HCV and in the development of screening assays for the identification of antiviral agents for HCV.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
ΑT	Austria	FR	France	LU	Luxembourg	SN	Senegal
ΑU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
ΑZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of Americ
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

- 1 -

•

Title Of Invention

Cloned Genomes Of Infectious Hepatitis C Viruses And Uses Thereof

5

This application claims the benefit of U.S. Provisional Application No. 60/053,062 filed July 18, 1997.

10

Field Of Invention

15

The present invention relates to molecular approaches to the production of nucleic acid sequences which comprise the genome of infectious hepatitis C viruses. In particular, the invention provides nucleic acid sequences which comprise the genomes of infectious hepatitis C viruses of genotype 1a and 1b strains. The invention therefore relates to the use of these sequences, and polypeptides encoded by all or part of these sequences, in the development of vaccines and diagnostic assays for HCV and in the development of screening assays for the identification of antiviral agents for HCV.

20

Background Of Invention

25

Hepatitis C virus (HCV) has a positive-sense single-strand RNA genome and is a member of the virus family Flaviviridae (Choo et al., 1991; Rice, 1996). As for all positive-stranded RNA viruses, the genome of HCV functions as mRNA from which all viral proteins necessary for propagation are translated.

30

The viral genome of HCV is approximately 9600 nucleotides (nts) and consists of a highly conserved 5' untranslated region (UTR), a single long open reading

- 2 -

frame (ORF) of approximately 9,000 nts and a complex 3'
UTR. The 5' UTR contains an internal ribosomal entry site
(Tsukiyama-Kohara et al., 1992; Honda et al., 1996). The

3' UTR consists of a short variable region, a polypyrimidine tract of variable length and, at the 3' end, a highly conserved region of approximately 100 nts (Kolykhalov et al., 1996; Tanaka et al., 1995; Tanaka et

al., 1996; Yamada et al., 1996). The last 46 nucleotides of this conserved region were predicted to form a stable stem-loop structure thought to be critical for viral replication (Blight and Rice, 1997; Ito and Lai, 1997; Tsuchihara et al., 1997). The ORF encodes a large

polypeptide precursor that is cleaved into at least 10 proteins by host and viral proteinases (Rice, 1996). The predicted envelope proteins contain several conserved N-linked glycosylation sites and cysteine residues (Okamoto et al., 1992a). The NS3 gene encodes a serine protease

and an RNA helicase and the NS5B gene encodes an RNA-dependent RNA polymerase.

Globally, six major HCV genotypes (genotypes 1-6) and multiple subtypes (a, b, c, etc.) have been identified (Bukh et al., 1993; Simmonds et al., 1993). The most divergent HCV isolates differ from each other by more than 30% over the entire genome (Okamoto et al., 1992a) and HCV circulates in an infected individual as a quasispecies of closely related genomes (Bukh et al., 1995; Farci et al., 1997).

At present, more than 80% of individuals infected with HCV become chronically infected and these chronically infected individuals have a relatively high

30

25

- 3 -

risk of developing chronic hepatitis, liver cirrhosis and hepatocellular carcinoma (Hoofnagle, 1997). In the U.S., HCV genotypes la and 1b constitute the majority of infections while in many other areas, especially in Europe and Japan, genotype 1b predominates.

The only effective therapy for chronic hepatitis C, interferon (IFN), induces a sustained response in less than 25% of treated patients (Fried and Hoofnagle, 1995). Consequently, HCV is currently the most common cause of end stage liver failure and the reason for about 30% of liver transplants performed in the U.S. (Hoofnagle, 1997). In addition, a number of recent studies suggested that the severity of liver disease and the outcome of therapy may be genotype-dependent (reviewed in Bukh et al., 1997). In particular, these studies suggested that infection with HCV genotype 1b was associated with more severe liver disease (Brechot, 1997) and a poorer response to IFN therapy (Fried and Hoofnagle, 1995). As a result of the inability to develop a universally effective therapy against HCV infection, it is estimated that there are still more than 25,000 new infections yearly in the U.S. (Alter 1997) Moreover, since there is no vaccine for HCV, HCV remains a serious public health problem.

However, despite the intense interest in the development of vaccines and therapies for HCV, progress has been hindered by the absence of a useful cell culture system and the lack of any small animal model for laboratory study. For example, while replication of HCV in several cell lines has been reported, such observations have turned out not to be highly reproducible. In

30

٥

5

10

15

20

- 4 -

addition, the chimpanzee is the only animal model, other than man, for this disease. Consequently, HCV has been able to be studied only by using clinical materials obtained from patients or experimentally infected chimpanzees (an animal model whose availability is very limited).

However, several researchers have recently reported the construction of infectious cDNA clones of HCV, the identification of which would permit a more effective search for susceptible cell lines and facilitate molecular analysis of the viral genes and their function. For example, Dash et al., (1997) and Yoo et al., (1995) reported that RNA transcripts from cDNA clones of HCV-1 (genotype 1a) and HCV-N (genotype 1b), respectively, resulted in viral replication after transfection into human hepatoma cell lines. Unfortunately, the viability of these clones was not tested in vivo and concerns were raised about the infectivity of these cDNA clones in vitro (Fausto, 1997). In addition, both clones did not contain the terminal 98 conserved nucleotides at the very 3' end of the UTR.

Kolykhalov et al., (1997) and Yanagi et al.

(1997) reported the derivation from HCV strain H77 (which is genotype 1a) of cDNA clones of HCV that are infectious for chimpanzees. However, while these infectious clones will aid in studying HCV replication and pathogenesis and will provide an important tool for development of in vitro replication and propagation systems, it is important to have infectious clones of more than one genotype given the extensive genetic heterogeneity of HCV and the potential

5

10

15

20

25

- 5 -

impact of such heterogeneity on the development of effective therapies and vaccines for HCV.

Summary Of The Invention

The present invention relates to nucleic acid sequences which comprise the genome of infectious hepatitis C viruses and in particular, nucleic acid sequences which comprise the genome of infectious hepatitis C viruses of genotype 1a and 1b strains. It is therefore an object of the invention to provide nucleic acid sequences which encode infectious hepatitis C viruses. Such nucleic acid sequences are referred to throughout the application as "infectious nucleic acid sequences".

For the purposes of this application, nucleic acid sequence refers to RNA, DNA, cDNA or any variant thereof capable of directing host organism synthesis of a hepatitis C virus polypeptide. It is understood that nucleic acid sequence encompasses nucleic acid sequences, which due to degeneracy, encode the same polypeptide sequence as the nucleic acid sequences described herein.

The invention also relates to the use of the infectious nucleic acid sequences to produce chimeric genomes consisting of portions of the open reading frames of infectious nucleic acid sequences of other genotypes (including, but not limited to, genotypes 1, 2, 3, 4, 5 and 6) and subtypes (including, but not limited to, subtypes 1a, 1b, 2a, 2b, 2c, 3a 4a-4f, 5a and 6a) of HCV. For example infectious nucleic acid sequence of the 1a and 1b strains H77 and HC-J4, respectively, described herein

30

0

5

10

15

20

- 6 -

5

10

15

can be used to produce chimeras with sequences from the genomes of other strains of HCV from different genotypes or subtypes. Nucleic acid sequences which comprise sequence from the open-reading frames of 2 or more HCV genotypes or subtypes are designated "chimeric nucleic acid sequences".

The invention further relates to mutations of the infectious nucleic acid sequences of the invention where mutation includes, but is not limited to, point mutations, deletions and insertions. In one embodiment, a gene or fragment thereof can be deleted to determine the effect of the deleted gene or genes on the properties of the encoded virus such as its virulence and its ability to replicate. In an alternative embodiment, a mutation may be introduced into the infectious nucleic acid sequences to examine the effect of the mutation on the properties of the virus in the host cell.

20 The invention also relates to the introduction of mutations or deletions into the infectious nucleic acid sequences in order to produce an attenuated hepatitis C virus suitable for vaccine development.

The invention further relates to the use of the infectious nucleic acid sequences to produce attenuated viruses via passage in vitro or in vivo of the viruses produced by transfection of a host cell with the infectious nucleic acid sequence.

The present invention also relates to the use of the nucleic acid sequences of the invention or fragments thereof in the production of polypeptides where "nucleic acid sequences of the invention" refers to infectious

35

30

- 7 -

0

5

nucleic acid sequences, mutations of infectious nucleic acid sequences, chimeric nucleic acid sequences and sequences which comprise the genome of attenuated viruses produced from the infectious nucleic acid sequences of the invention. The polypeptides of the invention, especially structural polypeptides, can serve as immunogens in the development of vaccines or as antigens in the development of diagnostic assays for detecting the presence of HCV in biological samples.

10

The invention therefore also relates to vaccines for use in immunizing mammals especially humans against hepatitis C. In one embodiment, the vaccine comprises one or more polypeptides made from a nucleic acid sequence of the invention or fragment thereof. In a second embodiment, the vaccine comprises a hepatitis C virus produced by transfection of host cells with the nucleic acid sequences of the invention.

20

25

15

The present invention therefore relates to methods for preventing hepatitis C in a mammal. In one embodiment the method comprises administering to a mammal a polypeptide or polypeptides encoded by a nucleic acid sequence of the invention in an amount effective to induce protective immunity to hepatitis C. In another embodiment, the method of prevention comprises administering to a mammal a hepatitis C virus of the invention in an amount effective to induce protective immunity against hepatitis C.

30

In yet another embodiment, the method of protection comprises administering to a mammal a nucleic acid sequence of the invention or a fragment thereof in an

- 8 -

amount effective to induce protective immunity against hepatitis C.

The invention also relates to hepatitis C viruses produced by host cells transfected with the nucleic acid sequences of the present invention.

The invention therefore also provides pharmaceutical compositions comprising the nucleic acid sequences of the invention and/or their encoded hepatitis C viruses. The invention further provides pharmaceutical compositions comprising polypeptides encoded by the nucleic acid sequences of the invention or fragments thereof. The pharmaceutical compositions of the invention may be used prophylactically or therapeutically.

The invention also relates to antibodies to the hepatitis C viruses of the invention or their encoded polypeptides and to pharmaceutical compositions comprising these antibodies.

The present invention further relates to polypeptides encoded by the nucleic acid sequences of the invention fragments thereof. In one embodiment, said polypeptide or polypeptides are fully or partially purified from hepatitis C virus produced by cells transfected with nucleic acid sequence of the invention. In another embodiment, the polypeptide or polypeptides are produced recombinantly from a fragment of the nucleic acid sequences of the invention. In yet another embodiment, the polypeptides are chemically synthesized.

The invention also relates to the use of the nucleic acid sequences of the invention to identify cell

5

10

15

20

25

- 9 -

lines capable of supporting the replication of HCV <u>in</u> <u>vitro</u>.

The invention further relates to the use of the nucleic acid sequences of the invention or their encoded proteases (e.g. NS3 protease) to develop screening assays to identify antiviral agents for HCV.

Brief Description Of Figures

Figure 1 shows a strategy for constructing fulllength cDNA clones of HCV strain H77. The long PCR products amplified with H1 and H9417R primers were cloned directly into pGEM-9zf(-) after digestion with Not I and Xba I (pH21, and pH50,). Next, the 3' UTR was cloned into both $pH21_I$ and $pH50_I$ after digestion with Afl II and Xba I (pH21 and pH50). pH21 was tested for infectivity in a chimpanzee. To improve the efficiency of cloning, we constructed a cassette vector with consensus 5' and 3' termini of H77. This cassette vector (pCV) was obtained by cutting out the BamHI fragment (nts 1358- 7530 of the H77 genome) from pH50, followed by religation. Finally, the long PCR products of H77 amplified with primers H1 and H9417R (H product) or primers A1 and H9417R (A product) were cloned into pCV after digestion with Age I and Afl II or with Pin AI and Bfr I. The latter procedure yielded multiple complete cDNA clones of strain H77 of HCV.

Figure 2 shows the results of gel

electrophoresis of long RT-PCR amplicons of the entire ORF of H77 and the transcription mixture of the infectious clone of H77. The complete ORF was amplified by long RT-PCR with the primers.H1 or A1 and H9417R from 105 GE of

0

5

10

15

20

- 10 -

0

5

10

H77. A total of 10 μ g of the consensus chimeric clone (pCV-H77C) linearized with Xba I was transcribed in a 100 μ l reaction with T7 RNA polymerase. Five μ l of the transcription mixture was analyzed by gel electrophoresis and the remainder of the mixture was injected into a chimpanzee. Lane 1, molecular weight marker; lane 2, products amplified with primers H1 and H9417R; lane 3, products amplified with primers A1 and H9417R; lane 4, transcription mixture containing the RNA transcripts and linearized clone pCV-H77C (12.5 kb).

of HCV strain H77 and the genetic heterogeneity of individual full-length clones compared with the consensus sequence of H77. Solid lines represent as changes.

Dashed lines represent silent mutations. A * in pH21 represents a point mutation at nt 58 in the 5' UTR. In the ORF, the consensus chimeric clone pCV-H77C had 11 nt differences [at positions 1625 (C→T), 2709 (T→C), 3380 (A→G), 3710 (C→T), 3914 (G→A), 4463 (T→C), 5058 (C→T),

one as change (F → L at as 790) compared with the

consensus sequence of H77. This clone was infectious.

Clone pH21 and pCV-H11 had 19 nts (7 as) and 64 nts (21
as) differences respectively, compared with the consensus sequence of H77. These two clones were not infectious. A

single point mutation in the 3' UTR at nucleotide 9406

 $(G\rightarrow A)$ introduced to create an Afl II cleavage site is not shown.

5834 (C \rightarrow T), 6734 (T \rightarrow C), 7154 (C \rightarrow T), and 7202 (T \rightarrow C)] and

- 11 -

0

Figures 4A-4F show the complete nucleotide sequence of a H77C clone produced according to the present invention and Figures 4G-4H show the amino acid sequence encoded by the H77C clone.

5

Figure 5 shows an agarose gel of long RT-PCR amplicons and transcription mixtures. Lanes 1 and 4: Molecular weight marker (Lambda/HindIII digest). Lanes 2 and 3: RT-PCR amplicons of the entire ORF of HC-J4. Lane 5: pCV-H77C transcription control (Yanagi et al., 1997). Lanes 6, 7, and 8: 1/40 of each transcription mixture of pCV-J4L2S, pCV-J4L4S and pCV-J4L6S, respectively, which was injected into the chimpanzee.

15

20

10

Figure 6 shows the strategy utilized for the construction of full-length cDNA clones of HCV strain HC-J4. The long PCR products were cloned as two separate fragments (L and S) into a cassette vector (pCV) with fixed 5' and 3' termini of HCV (Yanagi et al., 1997). Full-length cDNA clones of HC-J4 were obtained by inserting the L fragment from three pCV-J4L clones into three identical pCV-J4S9 clones after digestion with PinAI (isoschizomer of AgeI) and BfrI (isoschizomer of AfIII).

25

Figure 7 shows amino acid positions with a quasispecies of HC-J4 in the acute phase plasma pool obtained from an experimentally infected chimpanzee.

Cons-p9: consensus amino acid sequence deduced from analysis of nine L fragments and nine S fragments (see Fig. 6). Cons-D: consensus sequence derived from direct sequencing of the PCR product. A, B, C: groups of similar viral species. Dot: amino acid identical to that in Cons-

35

- 12 -

0

5

p9. Capital letter: amino acid different from that in Cons-p9. Cons-F: composite consensus amino acid sequence combining Cons-p9 and Cons-D. Boxed amino acid: different from that in Cons-F. Shaded amino acid: different from that in all species A sequences. An *: defective ORF due to a nucleotide deletion (clone L1, aa 1097) or insertion (clone L7, aa 2770). Diagonal lines: fragments used to construct the infectious clone.

10

Figure 8 shows comparisons (percent difference) of nucleotide (nts. 156 - 8935) and predicted amino acid sequences (aa 1 - 2864) of L clones (species A, B, and C, this study), HC-J4/91 (Okamoto et al., 1992b) and HC-J4/83 (Okamoto et al., 1992b). Differences among species A sequences and among species B sequences are shaded.

15

Figure 9 shows UPGMA ("unweighted pair group method with arithmetic mean") trees of HC-J4/91 (Okamoto et al., 1992b), HC-J4/83 (Okamoto et al., 1992b), two prototype strains of genotype 1b (HCV-J, Kato et al., 1990; HCV-BK, Takamizawa et al., 1991), and L clones (this study).

25

20

Figure 10 shows the alignment of the HVR1 and HVR2 amino acid sequences of the E2 sequences of nine L clones of HC-J4 (species A, B, and C) obtained from an early acute phase plasma pool of an experimentally infected chimpanzee compared with the sequences of eight clones (HC-J4/91-20 through HC-J4/91-27, Okamoto et al., 1992b) derived from the inoculum. Dot: an amino acid identical to that in the top line. Capital letters: amino acid different from that in the top line.

- 13 -

0

5

Figure 11 shows the alignment of the 5' UTR and the 3' UTR sequences of infectious clones of genotype la (pCV-H77C) and 1b (pCV-J4L6S). Top line: consensus sequence of the indicated strain. Dot: identity with consensus sequence. Capital letter: different from the consensus sequence. Dash: deletion. Underlined: PinAI and BfrI cleavage site. Numbering corresponds to the HCV sequence of pCV-J4L6S.

10

Figure 12 shows a comparison of individual full-length cDNA clones of the ORF of HCV strain HC-J4 with the consensus sequence (see Fig. 7). Solid lines: amino acid changes. Dashed lines: silent mutations. Clone pCV-J4L6S was infectious <u>in vivo</u> whereas clones pCV-J4L2S and pCV-J4L4S were not.

15

20

25

Figure 13 shows biochemical (ALT levels) and PCR analyses of a chimpanzee following percutaneous intrahepatic transfection with RNA transcripts of the infectious clone of pCV-J4L2S, pCV-J4L4S and pCV-J4L6S. The ALT serum enzyme levels were measured in units per liter (u/l). For the PCR analysis, "HCV RNA" represented by an open rectangle indicates a serum sample that was negative for HCV after nested PCR; "HCV RNA" represented by a closed rectangle indicates that the serum sample was positive for HCV and HCV GE titer on the right-hand y-axis represents genome equivalents.

30

Figures 14A-14F show the nucleotide sequence of the infectious clone of genotype 1b strain HC-J4 and Figures 14G-14H show the amino acid sequence encoded by the HC-J4 clone.

- 14 -

0

5

10

15

20

25

Figure 15 shows the strategy for constructing a chimeric HCV clone designated pH77CV-J4 which contains the nonstructural region of the infectious clone of genotype 1a strain H77 and the structural region of the infectious clone of genotype 1b strain HC-J4.

Figures 16A-16F show the nucleotide sequence of the chimeric 1a/lb clone pH77CV-J4 of Figure 15 and Figures 16G-16H show the amino acid sequence encoded by the chimeric 1a/lb clone.

Figures 17A and 17B show the sequence of the 3' untranslated region remaining in various 3' deletion mutants of the 1a infectious clone pCV-H77C and the strategy utilized in constructing each 3' deletion mutant (Figures 17C-17G).

Of the seven deletion mutants shown, two (pCV-H77C(-98X) and (-42X)) have been constructed and tested for infectivity in chimpanzees (see Figures 17A and 17C) and the other six are to be constructed and tested for infectivity as described in Figures 17D-17G.

Figures 18A and 18B show biochemical (ALT levels), PCR (HCV RNA and HCV GE titer), serological (anti-HCV) and histopathological (Fig. 18B only) analyses of chimpanzees 1494 (Fig. 18A) and 1530 (Fig. 18B) following transfection with the infectious cDNA clone pCV-H77C.

The ALT serum enzyme levels were measured in units per ml (u/l). For the PCR analysis, "HCV RNA" represented by an open rectangle indicates a serum sample that was negative for HCV after nested PCR; "HCV RNA" represented by a closed rectangle indicates that the serum

PCT/US98/14688 WO 99/04008

- 15 -

sample was positive for HCV; and HCV GE titer on the right-hand y-axis represents genome equivalents.

The bar marked "anti-HCV" indicates samples that were positive for anti-HCV antibodies as determined by commercial assays. The histopathology scores in Figure 18B correspond to no histopathology (O), mild hepatitis (\mathcal{Q}) and moderate to severe hepatitis (\circ) .

DESCRIPTION OF THE INVENTION

10

The present invention relates to nucleic acid sequences which comprise the genome of an infectious hepatitis C virus. More specifically, the invention relates to nucleic acid sequences which encode infectious hepatitis C viruses of genotype la and lb strains. embodiment, the infectious nucleic acid sequence of the invention has the sequence shown in Figures 4A-4F of this In another embodiment, the infectious application. nucleic acid sequence has the sequence shown in Figures 14A-14F and is contained in a plasmid construct deposited with the American Type Culture Collection (ATCC) on January 26, 1998 and having ATCC accession number 209596.

The invention also relates to "chimeric nucleic acid sequences" where the chimeric nucleic acid sequences consist of open-reading frame sequences taken from infectious nucleic acid sequences of hepatitis C viruses of different genotypes or subtypes.

In one embodiment, the chimeric nucleic acid sequence consists of sequence from the genome of an HCV strain belonging to one genotype or subtype which encodes structural polypeptides and sequence of an HCV strain

35

30

0

5

15

20

- 16 -

0

5

10

15

20

25

30

belonging to another genotype strain or subtype which encodes nonstructural polypeptides. Such chimeras can be produced by standard techniques of restriction digestion, PCR amplification and subcloning known to those of ordinary skill in the art.

In a preferred embodiment, the sequence encoding nonstructural polypeptides is from an infectious nucleic acid sequence encoding a genotype 1a strain where the construction of a chimeric 1a/1b nucleic acid sequence is described in Example 9 and the chimeric 1a/1b nucleic acid sequence is shown in Figures 16A-16F. It is believed that the construction of such chimeric nucleic acid sequences will be of importance in studying the growth and virulence properties of hepatitis C virus and in the production of hepatitis C viruses suitable to confer protection against multiple genotypes of HCV. For example, one might produce a "multivalent" vaccine by putting epitopes from several genotypes or subtypes into one clone. Alternatively one might replace just a single gene from an infectious sequence with the corresponding gene from the genomic sequence of a strain from another genotype or subtype or create a chimeric gene which contains portions of a gene from two genotypes or subtypes. Examples of genes which could be replaced or which could be made chimeric, include, but are not limited to, the E1, E2 and NS4 genes.

The invention further relates to mutations of the infectious nucleic acid sequences where "mutations" includes, but is not limited to, point mutations, deletions and insertions. Of course, one of ordinary skill in the art would recognize that the size of the

PCT/US98/14688

- 17 -

insertions would be limited by the ability of the resultant nucleic acid sequence to be properly packaged within the virion. Such mutation could be produced by techniques known to those of skill in the art such as site-directed mutagenesis, fusion PCR, and restriction digestion followed by religation.

In one embodiment, mutagenesis might be undertaken to determine sequences that are important for viral properties such as replication or virulence. example, one may introduce a mutation into the infectious nucleic acid sequence which eliminates the cleavage site between the NS4A and NS4B polypeptides to examine the effects on viral replication and processing of the polypeptide. Alternatively, one or more of the 3 amino acids encoded by the infectious 1b nucleic acid sequence shown in Figures 14A-14F which differ from the HC-J4 consensus sequence may be back mutated to the corresponding amino acid in the HC-J4 consensus sequence to determine the importance of these three amino acid changes to infectivity or virulence. In yet another embodiment, one or more of the amino acids from the noninfectious 1b clones pCV-J4L2S and pCV-J4L4S which differ from the consensus sequence may be introduced into the infectious 1b sequence shown in Figures 14A-14F.

In yet another example, one may delete all or part of a gene or of the 5' or 3' nontranslated region contained in an infectious nucleic acid sequence and then transfect a host cell (animal or cell culture) with the mutated sequence and measure viral replication in the host by methods known in the art such as RT-PCR. Preferred

35

30

WO 99/04008

5

10

15

20

- 18 -

0

5

genes include, but are not limited to, the P7, NS4B and NS5A genes. Of course, those of ordinary skill in the art will understand that deletion of part of a gene, preferably the central portion of the gene, may be preferable to deletion of the entire gene in order to conserve the cleavage site boundaries which exist between proteins in the HCV polyprotein and which are necessary for proper processing of the polyprotein.

10

15

In the alternative, if the transfection is into a host animal such as a chimpanzee, one can monitor the virulence phenotype of the virus produced by transfection of the mutated infectious nucleic acid sequence by methods known in the art such as measurement of liver enzyme levels (alanine aminotransferase (ALT) or isocitrate dehydrogenase (ICD)) or by histopathology of liver biopsies. Thus, mutations of the infectious nucleic acid sequences may be useful in the production of attenuated HCV strains suitable for vaccine use.

20

The invention also relates to the use of the infectious nucleic acid sequences of the present invention to produce attenuated viral strains via passage <u>in vitro</u> or <u>in vivo</u> of the virus produced by transfection with the infectious nucleic acid sequences.

25

The present invention therefore relates to the use of the nucleic acid sequences of the invention to identify cell lines capable of supporting the replication of HCV.

30

In particular, it is contemplated that the mutations of the infectious nucleic acid sequences of the invention and the production of chimeric sequences as

- 19 -

discussed above may be useful in identifying sequences critical for cell culture adaptation of HCV and hence, may be useful in identifying cell lines capable of supporting HCV replication.

Transfection of tissue culture cells with the nucleic acid sequences of the invention may be done by methods of transfection known in the art such as electroporation, precipitation with DEAE-Dextran or calcium phosphate or liposomes.

In one such embodiment, the method comprises the growing of animal cells, especially human cells, in vitro and transfecting the cells with the nucleic acid of the invention, then determining if the cells show indicia of HCV infection. Such indicia include the detection of viral antigens in the cell, for example, by immunofluorescent procedures well known in the art; the detection of viral polypeptides by Western blotting using antibodies specific therefor; and the detection of newly transcribed viral RNA within the cells via methods such as RT-PCR. The presence of live, infectious virus particles following such tests may also be shown by injection of cell culture medium or cell lysates into healthy, susceptible animals, with subsequent exhibition of the symptoms of HCV infection.

Suitable cells or cell lines for culturing HCV include, but are not limited to, lymphocyte and hepatocyte cell lines known in the art.

Alternatively, primary hepatocytes can be cultured, and then infected with HCV; or, the hepatocyte cultures could be derived from the livers of infected

30

5

10

15

20

101/05/0/140

- 20 -

0

5

10

15

20

25

30

chimpanzees. In addition, various immortalization methods known to those of ordinary skill in the art can be used to obtain cell-lines derived from hepatocyte cultures. For example, primary hepatocyte cultures may be fused to a variety of cells to maintain stability.

The present invention further relates to the <u>in</u>

<u>vitro</u> and <u>in vivo</u> production of hepatitis C viruses from

the nucleic acid sequences of the invention.

In one embodiment, the sequences of the invention can be inserted into an expression vector that functions in eukaryotic cells. Eukaryotic expression vectors suitable for producing high efficiency gene transfer in vivo are well known to those of ordinary skill in the art and include, but are not limited to, plasmids, vaccinia viruses, retroviruses, adenoviruses and adeno-associated viruses.

In another embodiment, the sequences contained in the recombinant expression vector can be transcribed in vitro by methods known to those of ordinary skill in the art in order to produce RNA transcripts which encode the hepatitis C viruses of the invention. The hepatitis C viruses of the invention may then be produced by transfecting cells by methods known to those of ordinary skill in the art with either the in vitro transcription mixture containing the RNA transcripts (see Example 4) or with the recombinant expression vectors containing the nucleic acid sequences described herein.

The present invention also relates to the construction of cassette vectors useful in the cloning of viral genomes wherein said vectors comprise a nucleic acid

5

10

15

20

25

sequence to be cloned, and said vector reading in the correct phase for the expression of the viral nucleic acid to be cloned. Such a cassette vector will, of course, also possess a promoter sequence, advantageously placed upstream of the sequence to be expressed. Cassette vectors according to the present invention are constructed according to the procedure described in Figure 1, for example, starting with plasmid pCV. Of course, the DNA to be inserted into said cassette vector can be derived from any virus, advantageously from HCV, and most advantageously from the H77 strain of HCV. The nucleic acid to be inserted according to the present invention can, for example, contain one or more open reading frames of the virus, for example, HCV. The cassette vectors of the present invention may also contain, optionally, one or more expressible marker genes for expression as an indication of successful transfection and expression of the nucleic acid sequences of the vector. To insure expression, the cassette vectors of the present invention will contain a promoter sequence for binding of the appropriate cellular RNA polymerase, which will depend on the cell into which the vector has been introduced. example, if the host cell is a bacterial cell, then said promoter will be a bacterial promoter sequence to which the bacterial RNA polymerases will bind.

The hepatitis C viruses produced from the sequences of the invention may be purified or partially purified from the transfected cells by methods known to those of ordinary skill in the art. In a preferred embodiment, the viruses are partially purified prior to

- 22 -

0

5

10

15

20

25

30

their use as immunogens in the pharmaceutical compositions and vaccines of the present invention.

The present invention therefore relates to the use of the hepatitis C viruses produced from the nucleic acid sequences of the invention as immunogens in live or killed (e.g., formalin inactivated) vaccines to prevent hepatitis C in a mammal.

In an alternative embodiment, the immunogen of the present invention may be an infectious nucleic acid sequence, a chimeric nucleic acid sequence, or a mutated infectious nucleic acid sequence which encodes a hepatitis C virus. Where the sequence is a cDNA sequence, the cDNAs and their RNA transcripts may be used to transfect a mammal by direct injection into the liver tissue of the mammal as described in the Examples.

Alternatively, direct gene transfer may be accomplished via administration of a eukaryotic expression vector containing a nucleic acid sequence of the invention.

In yet another embodiment, the immunogen may be a polypeptide encoded by the nucleic acid sequences of the invention. The present invention therefore also relates to polypeptides produced from the nucleic acid sequences of the invention or fragments thereof. In one embodiment, polypeptides of the present invention can be recombinantly produced by synthesis from the nucleic acid sequences of the invention or isolated fragments thereof, and purified, or partially purified, from transfected cells using methods already known in the art. In an alternative embodiment, the polypeptides may be purified or partially

- 23 -

purified from viral particles produced via transfection of a host cell with the nucleic acid sequences of the invention. Such polypeptides might, for example, include either capsid or envelope polypeptides prepared from the sequences of the present invention.

When used as immunogens, the nucleic acid sequences of the invention, or the polypeptides or viruses produced therefrom, are preferably partially purified prior to use as immunogens in pharmaceutical compositions and vaccines of the present invention. When used as a vaccine, the sequences and the polypeptide and virus products thereof, can be administered alone or in a suitable diluent, including, but not limited to, water, saline, or some type of buffered medium. The vaccine according to the present invention may be administered to an animal, especially a mammal, and most especially a human, by a variety of routes, including, but not limited to, intradermally, intramuscularly, subcutaneously, or in any combination thereof.

Suitable amounts of material to administer for prophylactic and therapeutic purposes will vary depending on the route selected and the immunogen (nucleic acid, virus, polypeptide) administered. One skilled in the art will appreciate that the amounts to be administered for any particular treatment protocol can be readily determined without undue experimentation. The vaccines of the present invention may be administered once or periodically until a suitable titer of anti-HCV antibodies appear in the blood. For an immunogen consisting of a nucleic acid sequence, a suitable amount of nucleic acid

30

5

10

15

20

0

5

sequence to be used for prophylactic purposes might be expected to fall in the range of from about 100 μg to about 5 mg and most preferably in the range of from about 500 μg to about 2mg. For a polypeptide, a suitable amount to use for prophylactic purposes is preferably 100 ng to 100 μg and for a virus 10^2 to 10^6 infectious doses. Such administration will, of course, occur prior to any sign of HCV infection.

10

15

20

25

A vaccine of the present invention may be employed in such forms as capsules, liquid solutions, suspensions or elixirs for oral administration, or sterile liquid forms such as solutions or suspensions. Any inert carrier is preferably used, such as saline or phosphatebuffered saline, or any such carrier in which the HCV of the present invention can be suitably suspended. vaccines may be in the form of single dose preparations or in multi-dose flasks which can be utilized for massvaccination programs of both animals and humans. purposes of using the vaccines of the present invention reference is made to Remington's Pharmaceutical Sciences, Mack Publishing Co., Easton, Pa., Osol (Ed.) (1980); and New Trends and Developments in Vaccines, Voller et al. (Eds.), University Park Press, Baltimore, Md. (1978), both of which provide much useful information for preparing and using vaccines. Of course, the polypeptides of the present invention, when used as vaccines, can include, as part of the composition or emulsion, a suitable adjuvant, such as alum (or aluminum hydroxide) when humans are to be vaccinated, to further stimulate production of antibodies by immune cells. When nucleic acids or viruses are used

35

0

for vaccination purposes, other specific adjuvants such as CpG motifs (Krieg, A.K. et al.(1995) and (1996)), may prove useful.

When the nucleic acids, viruses and polypeptides 5 of the present invention are used as vaccines or inocula, they will normally exist as physically discrete units suitable as a unitary dosage for animals, especially mammals, and most especially humans, wherein each unit will contain a predetermined quantity of active material 10 calculated to produce the desired immunogenic effect in association with the required diluent. The dose of said vaccine or inoculum according to the present invention is administered at least once. In order to increase the 15 antibody level, a second or booster dose may be administered at some time after the initial dose. need for, and timing of, such booster dose will, of course, be determined within the sound judgment of the administrator of such vaccine or inoculum and according to 20 sound principles well known in the art. For example, such booster dose could reasonably be expected to be advantageous at some time between about 2 weeks to about 6 months following the initial vaccination. Subsequent 25 doses may be administered as indicated.

The nucleic acid sequences, viruses and polypeptides of the present invention can also be administered for purposes of therapy, where a mammal, especially a primate, and most especially a human, is already infected, as shown by well known diagnostic measures. When the nucleic acid sequences, viruses or polypeptides of the present invention are used for such

- 26 -

0

5

10

15

20

25

therapeutic purposes, much of the same criteria will apply as when it is used as a vaccine, except that inoculation will occur post-infection. Thus, when the nucleic acid sequences, viruses or polypeptides of the present invention are used as therapeutic agents in the treatment of infection, the therapeutic agent comprises a pharmaceutical composition containing a sufficient amount of said nucleic acid sequences, viruses or polypeptides so as to elicit a therapeutically effective response in the organism to be treated. Of course, the amount of pharmaceutical composition to be administered will, as for vaccines, vary depending on the immunogen contained therein (nucleic acid, polypeptide, virus) and on the route of administration.

The therapeutic agent according to the present invention can thus be administered by, subcutaneous, intramuscular or intradermal routes. One skilled in the art will certainly appreciate that the amounts to be administered for any particular treatment protocol can be readily determined without undue experimentation. Of course, the actual amounts will vary depending on the route of administration as well as the sex, age, and clinical status of the subject which, in the case of human patients, is to be determined with the sound judgment of the clinician.

The therapeutic agent of the present invention

can be employed in such forms as capsules, liquid solutions, suspensions or elixirs, or sterile liquid forms such as solutions or suspensions. Any inert carrier is preferably used, such as saline, phosphate-buffered

- 27 -

saline, or any such carrier in which the HCV of the present invention can be suitably suspended. The therapeutic agents may be in the form of single dose preparations or in the multi-dose flasks which can be utilized for mass-treatment programs of both animals and humans. Of course, when the nucleic acid sequences, viruses or polypeptides of the present invention are used as therapeutic agents they may be administered as a single dose or as a series of doses, depending on the situation as determined by the person conducting the treatment.

The nucleic acids, polypeptides and viruses of the present invention can also be utilized in the production of antibodies against HCV. The term "antibody" is herein used to refer to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules. Examples of antibody molecules are intact immunoglobulin molecules, substantially intact immunoglobulin molecules and portions of an immunoglobulin molecule, including those portions known in the art as Fab, F(ab')₂ and F(v) as well as chimeric antibody molecules.

Thus, the polypeptides, viruses and nucleic acid sequences of the present invention can be used in the generation of antibodies that immunoreact (i.e., specific binding between an antigenic determinant-containing molecule and a molecule containing an antibody combining site such as a whole antibody molecule or an active portion thereof) with antigenic determinants on the surface of hepatitis C virus particles.

30

o

5

10

15

20

- 28 -

C

5

10

The present invention therefore also relates to antibodies produced following immunization with the nucleic acid sequences, viruses or polypeptides of the present invention. These antibodies are typically produced by immunizing a mammal with an immunogen or vaccine to induce antibody molecules having immunospecificity for polypeptides or viruses produced in response to infection with the nucleic acid sequences of the present invention. When used in generating such antibodies, the nucleic acid sequences, viruses, or polypeptides of the present invention may be linked to some type of carrier molecule. The resulting antibody molecules are then collected from said mammal. Antibodies produced according to the present invention have the unique advantage of being generated in response to authentic, functional polypeptides produced according to the actual cloned HCV genome.

20

15

The antibody molecules of the present invention may be polyclonal or monoclonal. Monoclonal antibodies are readily produced by methods well known in the art. Portions of immunoglobin molecules, such as Fabs, as well as chimeric antibodies, may also be produced by methods well known to those of ordinary skill in the art of generating such antibodies.

25

30

The antibodies according to the present invention may also be contained in blood plasma, serum, hybridoma supernatants, and the like. Alternatively, the antibody of the present invention is isolated to the extent desired by well known techniques such as, for example, using DEAE Sephadex. The antibodies produced

PCT/US98/14688

٥

5

10

15

20

25

WO 99/04008

according to the present invention may be further purified so as to obtain specific classes or subclasses of antibody such as IgM, IgG, IgA, and the like. Antibodies of the IgG class are preferred for purposes of passive protection.

The antibodies of the present invention are useful in the prevention and treatment of diseases caused by hepatitis C virus in animals, especially mammals, and most especially humans.

In providing the antibodies of the present invention to a recipient mammal, preferably a human, the dosage of administered antibodies will vary depending on such factors as the mammal's age, weight, height, sex, general medical condition, previous medical history, and the like.

In general, it will be advantageous to provide the recipient mammal with a dosage of antibodies in the range of from about 1 mg/kg body weight to about 10 mg/kg body weight of the mammal, although a lower or higher dose may be administered if found desirable. Such antibodies will normally be administered by intravenous or intramuscular route as an inoculum. The antibodies of the present invention are intended to be provided to the recipient subject in an amount sufficient to prevent, lessen or attenuate the severity, extent or duration of any existing infection.

The antibodies prepared by use of the nucleic acid sequences, viruses or polypeptides of the present invention are also highly useful for diagnostic purposes. For example, the antibodies can be used as in vitro

35

- 30 -

0

5

10

15

20

25

30

in the art.

diagnostic agents to test for the presence of HCV in biological samples taken from animals, especially humans. Such assays include, but are not limited to, radioimmunoassays, EIA, fluorescence, Western blot analysis and ELISAs. In one such embodiment, the biological sample is contacted with antibodies of the present invention and a labeled second antibody is used to detect the presence of HCV to which the antibodies are bound.

Such assays may be, for example, a direct protocol (where the labeled first antibody is immunoreactive with the antigen, such as, for example, a polypeptide on the surface of the virus), an indirect protocol (where a labeled second antibody is reactive with the first antibody), a competitive protocol (such as would involve the addition of a labeled antigen), or a sandwich protocol (where both labeled and unlabeled antibody are used), as well as other protocols well known and described

In one embodiment, an immunoassay method would utilize an antibody specific for HCV envelope determinants and would further comprise the steps of contacting a biological sample with the HCV-specific antibody and then detecting the presence of HCV material in the test sample using one of the types of assay protocols as described above. Polypeptides and antibodies produced according to the present invention may also be supplied in the form of a kit, either present in vials as purified material, or present in compositions and suspended in suitable diluents as previously described.

- 31 -

0

5

10

15

20

25

In a preferred embodiment, such a diagnostic test kit for detection of HCV antigens in a test sample comprises in combination a series of containers, each container a reagent needed for such assay. Thus, one such container would contain a specific amount of HCV-specific antibody as already described, a second container would contain a diluent for suspension of the sample to be tested, a third container would contain a positive control and an additional container would contain a negative control. An additional container could contain a blank.

For all prophylactic, therapeutic and diagnostic uses, the antibodies of the invention and other reagents, plus appropriate devices and accessories, may be provided in the form of a kit so as to facilitate ready availability and ease of use.

The present invention also relates to the use of nucleic acid sequences and polypeptides of the present invention to screen potential antiviral agents for antiviral activity against HCV. Such screening methods are known by those of skill in the art. Generally, the antiviral agents are tested at a variety of concentrations, for their effect on preventing viral replication in cell culture systems which support viral replication, and then for an inhibition of infectivity or of viral pathogenicity (and a low level of toxicity) in an animal model system.

In one embodiment, animal cells (especially human cells) transfected with the nucleic acid sequences of the invention are cultured <u>in vitro</u> and the cells are treated with a candidate antiviral agent (a chemical,

35

- 32 -

0

5

10

15

20

25

peptide etc.) for antiviral activity by adding the candidate agent to the medium. The treated cells are then exposed, possibly under transfecting or fusing conditions known in the art, to the nucleic acid sequences of the present invention. A sufficient period of time would then be allowed to pass for infection to occur, following which the presence or absence of viral replication would be determined versus untreated control cells by methods known to those of ordinary skill in the art. Such methods include, but are not limited to, the detection of viral antigens in the cell, for example, by immunofluorescent procedures well known in the art; the detection of viral polypeptides by Western blotting using antibodies specific therefor; the detection of newly transcribed viral RNA within the cells by RT-PCR; and the detection of the presence of live, infectious virus particles by injection of cell culture medium or cell lysates into healthy, susceptible animals, with subsequent exhibition of the symptoms of HCV infection. A comparison of results obtained for control cells (treated only with nucleic acid sequence) with those obtained for treated cells (nucleic acid sequence and antiviral agent) would indicate, the degree, if any, of antiviral activity of the candidate antiviral agent. Of course, one of ordinary skill in the art would readily understand that such cells can be treated with the candidate antiviral agent either before or after exposure to the nucleic acid sequence of the present invention so as to determine what stage, or stages, of viral infection and replication said agent is effective against.

35

PCT/US98/14688

0

5

In an alternative embodiment, a protease such as NS3 protease produced from a nucleic acid sequence of the invention may be used to screen for protease inhibitors which may act as antiviral agents. The structural and nonstructural regions of the HCV genome, including nucleotide and amino acid locations, have been determined, for example, as depicted in Houghton, M. (1996), Fig. 1; and Major, M.E. et al. (1997), Table 1.

10

15

20

Such above-mentioned protease inhibitors may take the form of chemical compounds or peptides which mimic the known cleavage sites of the protease and may be screened using methods known to those of skill in the art (Houghton, M. (1996) and Major, M.E. et al. (1997)). For example, a substrate may be employed which mimics the protease's natural substrate, but which provides a detectable signal (e.g. by fluorimetric or colorimetric methods) when cleaved. This substrate is then incubated with the protease and the candidate protease inhibitor under conditions of suitable pH, temperature etc. to detect protease activity. The proteolytic activities of the protease in the presence or absence of the candidate inhibitor are then determined.

25

In yet another embodiment, a candidate antiviral agent (such as a protease inhibitor) may be directly assayed <u>in vivo</u> for antiviral activity by administering the candidate antiviral agent to a chimpanzee transfected with a nucleic acid sequence of the invention and then measuring viral replication <u>in vivo</u> via methods such as RT-PCR. Of course, the chimpanzee may be treated with the candidate agent either before or after transfection with

35

- 34 -

0

the infectious nucleic acid sequence so as to determine what stage, or stages, of viral infection and replication the agent is effective against.

5

The invention also provides that the nucleic acid sequences, viruses and polypeptides of the invention may be supplied in the form of a kit, alone or in the form of a pharmaceutical composition.

10

All scientific publication and/or patents cited herein are specifically incorporated by reference. The following examples illustrate various aspects of the invention but are in no way intended to limit the scope thereof.

15

EXAMPLES

MATERIALS AND METHODS For Examples 1-4

Collection of Virus

20

25

30

Hepatitis C virus was collected and used as a source for the RNA used in generating the cDNA clones according to the present invention. Plasma containing strain H77 of HCV was obtained from a patient in the acute phase of transfusion-associated non-A, non-B hepatitis (Feinstone et al (1981)). Strain H77 belongs to genotype la of HCV (Ogata et al (1991), Inchauspe et al (1991)). The consensus sequence for most of its genome has been determined (Kolyakov et al (1996), Ogata et al (1991), Inchauspe et al (1991) and Farci et al (1996)).

5

10

15

- 35 -

RNA Purification

Viral RNA was collected and purified by conventional means. In general, total RNA from 10 μ l of H77 plasma was extracted with the TRIzol system (GIBCO BRL). The RNA pellet was resuspended in 100 μ l of 10 mM dithiothreitol (DTT) with 5% (vol/vol) RNasin (20 - 40 units/ μ l) (available from Promega) and 10 μ l aliquots were stored at -80°C. In subsequent experiments RT-PCR was performed on RNA equivalent to 1 μ l of H77 plasma, which contained an estimated 10⁵ genome equivalents (GE) of HCV (Yanaqi et al (1996)).

Primers used in the RT-PCR process were deduced from the genomic sequences of strain H77 according to procedures already known in the art (see above) or else were determined specifically for use herein. The primers generated for this purpose are listed in Table 1.

20

25

30

7O 59/04006 FC1/OS96/14

- 36 -

0

Table 1. Oligonucleotides used for PCR amplification of strain H77 of HCV

	Designation Sequence (5' → 3')*
H9261F	GGCTACAGCGGGGGAGACATTTATCACAGC
H3'X58R	TCATGCGGCTCACGGACCTTTCACAGCTAG
H9282F	GTCCAAGCTTATCACAGCGTGTCTCATGCCCGGCCCCG
H3'X45R	CGTCTCTAGAGGACCTTTCACAGCTAGCCGTGACTAGGG
H9375F	TGAAGGTTGGGGTAAACACTCCGGCCTCTTAGGCCATT
H3'X-35R	ACATGATCTGCAGAGAGGCCAGTATCAGCACTCTC
H9386F	GTCCAAGCTTACGCGTAAACACTCCGGCCTCCTTAAGCCATTCCTG
H3'X-38R	CGTCTCTAGACATGATCTGCAGAGAGGCCAGTATCAGCACTCTCTGC
H1	TTTTTTTGCGGCCGCTAATACGACTCACTATAGCCAGCCCCCTGAT-
	GGGGGCGACACTCCACCATG
A1	ACTGTCTTCACGCAGAAAGCGTCTAGCCAT
H9417R	CGTCTCTAGACAGGAAATGGCTTAAGAGGCCGGAGTGTTTACC
* HC	V sequences are shown in plain text, non-HCV-specific
sequences a	are shown in boldface and artificial cleavage sites
	ONA cloning are underlined. The core sequenceof the
T7 promoter	in primer H1 is shown in italics.
Prime	ers for long RT-PCR were size-purified.
	H3'X58R H9282F H3'X45R H9375F H3'X-35R H9386F H3'X-38R H1 A1 H9417R * H0 sequences a used for ci

cDNA Synthesis

The RNA was denatured at 65°C for 2 min, and cDNA synthesis was performed in a 20 μ l reaction volume with Superscript II reverse transcriptase (from GIBCO/BRL) at 42 °C for 1 hour using specific antisense primers as described previously (Tellier et al (1996)). The cDNA mixture was treated with RNase H and RNase T1 (GIBCO/BRL) for 20 min at 37 °C.

Amplification and Cloning of the 3' UTR

The 3' UTR of strain H77 was amplified by PCR in two different assays. In both of these nested PCR reactions the first round of PCR was performed in a total volume of 50 μ l in 1 x buffer, 250 μ mol of each deoxynucleoside triphosphate (dNTP; Pharmacia), 20 pmol

35

20

each of external sense and antisense primers, 1 μ l of the Advantage KlenTaq polymerase mix (from Clontech) and 2 μ l of the final cDNA reaction mixture. In the second round of PCR, 5 μ l of the first round PCR mixture was added to 45 μ l of PCR mixture prepared as already described. 5 round of PCR (35 cycles), which was performed in a Perkin Elmer DNA thermal cycler 480, consisted of denaturation at 94 °C for 1 min (in 1st cycle 1 min 30 sec), annealing at 60°C for 1 min and elongation at 68°C for 2 min. 10 experiment a region from NS5B to the conserved region of the 3' UTR was amplified with the external primers H9261F and H3'X58R, and the internal primers H9282F and H3'X45R (Table 1). In another experiment, a segment of the 15 variable region to the very end of the 3' UTR was amplified with the external primers H9375F and H3'X-35R, and the internal primers H9386F and H3'X-38R (Table 1, Fig. 1). Amplified products were purified with QIAquick 20 PCR purification kit (from QIAGEN), digested with Hind III and Xba I (from Promega), purified by either gel electrophoresis or phenol/chloroform extraction, and then cloned into the multiple cloning site of plasmid pGEM-9zf(-) (Promega) or pUC19 (Pharmacia). Cloning of cDNA 25 into the vector was performed with T4 DNA ligase (Promega) by standard procedures.

Amplification of Near Full-Length H77 Genomes by Long PCR

The reactions were performed in a total volume of 50 μ l in 1 x buffer, 250 μ mol of each dNTP, 10 pmol each of sense and antisense primers, 1 μ l of the Advantage

- 38 -

0

5

KlenTaq polymerase mix and 2 μ l of the cDNA reaction mixture (Tellier et al (1996)). A single PCR round of 35 cycles was performed in a Robocycler thermal cycler (from Stratagene), and consisted of denaturation at 99 °C for 35 sec, annealing at 67 °C for 30 sec and elongation at 68 °C for 10 min during the first 5 cycles, 11 min during the next 10 cycles, 12 min during the following 10 cycles and 13 min during the last 10 cycles. To amplify the complete ORF of HCV by long RT-PCR we used the sense primers H1 or A1 deduced from the 5' UTR and the antisense primer H9417R deduced from the variable region of the 3' UTR (Table 1, Fig. 1).

15

10

Construction of Full-Length H77 cDNA Clones

The long PCR products amplified with H1 and H9417R primers were cloned directly into pGEM-9zf(-) after digestion with Not I and Xba I (from Promega) (as per 20 Fig. 1). Two clones were obtained with inserts of the expected size, pH21, and pH50,. Next, the chosen 3' UTR was cloned into both pH21, and pH50, after digestion with Afl II and Xba I (New England Biolabs). DH5α competent 25 cells (GIBCO/BRL) were transformed and selected with LB agar plates containing 100 μ g/ml ampicillin (from SIGMA). Then the selected colonies were cultured in LB liquid containing ampicillin at 30°C for ~18-20 hrs 30 (transformants containing full-length or near full-length cDNA of H77 produced a very low yield of plasmid when cultured at 37 °C or for more than 24 hrs). After small scale preparation (Wizard Plus Minipreps DNA Purification

15

20

Systems, Promega) each plasmid was retransformed to select a single clone, and large scale preparation of plasmid DNA was performed with a QIAGEN plasmid Maxi kit.

5 Cloning of Long RT-PCR Products Into a Cassette Vector

with consensus 5' and 3' termini of HCV strain H77 was constructed (Fig. 1). This cassette vector (pCV) was obtained by cutting out the BamHI fragment (nts 1358 - 7530 of the H77 genome) from pH50, followed by religation. Next, the long PCR products of H77 amplified with H1 and H9417R or A1 and H9417R primers were purified (Geneclean spin kit; BIO 101) and cloned into pCV after digestion with Age I and Afl II(New England Biolabs) or with Pin AI (isoschizomer of Age I) and Bfr I (isoschizomer of Afl II) (Boehringer Mannheim). Large scale preparations of the plasmids containing full-length cDNA of H77 were performed as described above.

Construction of H77 Consensus Chimeric cDNA Clone

25 encoding the consensus amino acid sequence was constructed by making a chimera from four of the cDNA clones obtained above. This consensus chimera, pCV-H77C, was constructed in two ligation steps by using standard molecular procedures and convenient cleavage sites and involved first a two piece ligation and then a three piece ligation. Large scale preparation of pCV-H77C was performed as already described.

- 40 -

In Vitro Transcription

Plasmids containing the full-length HCV cDNA were linearized with Xba I (from Promega), and purified by phenol/chloroform extraction and ethanol precipitation. A 100 μ l reaction mixture containing 10 μ g of linearized plasmid DNA, 1 x transcription buffer, 1 mM ATP, CTP, GTP and UTP, 10mM DTT, 4% (v/v) RNasin (20-40 units/ μ l) and 2 μ l of T7 RNA polymerase (Promega) was incubated at 37 °C for 2 hrs. Five μ l of the reaction mixture was analyzed by agarose gel electrophoresis followed by ethidium bromide staining. The transcription reaction mixture was diluted with 400 μ l of ice-cold phosphate-buffered saline without calcium or magnesium, immediately frozen on dry ice and stored at -80 °C. The final nucleic acid mixture was injected into chimpanzees within 24 hrs.

Intrahepatic Transfection of Chimpanzees

Laparotomy was performed and aliquots from two transcription reactions were injected into 6 sites of the exposed liver (Emerson et al (1992). Serum samples were collected weekly from chimpanzees and monitored for liver enzyme levels and anti-HCV antibodies. Weekly samples of 100 µl of serum were tested for HCV RNA in a highly sensitive nested RT-PCR assay with AmpliTaq Gold (Perkin Elmer) (Yanagi et al (1996); Bukh et al (1992)). The genome titer of HCV was estimated by testing 10-fold serial dilutions of the extracted RNA in the RT-PCR assay (Yanagi et al (1996)). The two chimpanzees used in this

5

10

15

o

5

10

20

25

30

study were maintained under conditions that met all requirements for their use in an approved facility.

The consensus sequence of the complete ORF from HCV genomes recovered at week 2 post inoculation (p.i) was determined by direct sequencing of PCR products obtained in long RT-PCR with primers A1 and H9417R followed by nested PCR of 10 overlapping fragments. The consensus sequence of the variable region of the 3' UTR was determined by direct sequencing of an amplicon obtained in nested RT-PCR as described above. Finally, we amplified selected regions independently by nested RT-PCR with AmpliTaq Gold.

15 Sequence Analysis

Both strands of DNA from PCR products, as well as plasmids, were sequenced with the ABI PRISM Dye Termination Cycle Sequencing Ready Reaction Kit using Taq DNA polymerase (Perkin Elmer) and about 100 specific sense and antisense sequence primers.

The consensus sequence of HCV strain H77 was determined in two different ways. In one approach, overlapping PCR products were directly sequenced, and amplified in nested RT-PCR from the H77 plasma sample. The sequence analyzed (nucleotides (nts) 35-9417) included the entire genome except the very 5' and 3' termini. In the second approach, the consensus sequence of nts 157-9384 was deduced from the sequences of 18 full-length cDNA clones.

- 42 -

EXAMPLE 1

<u>Variability in the sequence of the 3' UTR of HCV strain</u> H77

5 The heterogeneity of the 3' UTR was analyzed by cloning and sequencing of DNA amplicons obtained in nested 19 clones containing sequences of the entire RT-PCR. variable region, the poly U-UC region and the adjacent 19 nt of the conserved region, and 65 clones containing 10 sequences of the entire poly U-UC region and the first 63 nts of the conserved region were analyzed. This analysis confirmed that the variable region consisted of 43 nts, including two conserved termination codons (Han et al 15 (1992)). The sequence of the variable region was highly conserved within H77 since only 3 point mutations were found among the 19 clones analyzed. A poly U-UC region was present in all 84 clones analyzed. However, its length varied from 71-141 nts. The length of the poly U 20 region was 9-103 nts, and that of the poly UC region was 35-85 nts. The number of C residues increased towards the 3' end of the poly UC region but the sequence of this region is not conserved. The first 63 nts of the 25 conserved region were highly conserved among the clones analyzed, with a total of only 14 point mutations. confirm the validity of the analysis, the 3' UTR was reamplified directly from a full-length cDNA clone of HCV 30 (see below) by the nested-PCR procedure with the primers in the variable region and at the very 3' end of the HCV genome and cloned the PCR product. Eight clones had 1-7 nt deletions in the poly U region. Furthermore, although

PCT/US98/14688

0

5

10

15

20

25

the C residues of the poly UC region were maintained, the spacing of these varied because of 1-2 nt deletions of U residues. These deletions must be artifacts introduced by PCR and such mistakes may have contributed to the heterogeneity originally observed in this region.

However, the conserved region of the 3' UTR was amplified correctly, suggesting that the deletions were due to

difficulties in transcribing a highly repetitive sequence.

One of the 3' UTR clones was selected for engineering of full-length cDNA clones of H77. This clone had the consensus variable sequence except for a single point mutation introduced to create an Afl II cleavage site, a poly U-UC stretch of 81 nts with the most commonly observed UC pattern and the consensus sequence of the complete conserved region of 101 nts, including the distal 38 nts which originated from the antisense primer used in the amplification. After linearization with Xba I, the DNA template of this clone had the authentic 3' end.

EXAMPLE 2

The Entire Open Reading Frame of H77 Amplified in One Round of Long RT-PCR

It had been previously demonstrated that a 9.25 kb fragment of the HCV genome from the 5' UTR to the 3' end of NS5B could be amplified from 10⁴ GE (genome equivalents) of H77 by a single round of long RT-PCR (Tellier et al (1996a)). In the current study, by optimizing primers and cycling conditions, the entire ORF of H77 was amplified in a single round of long RT-PCR with primers from the 5' UTR and the variable region of the 3'

U 99/04008 FC1/0598/140

- 44 -

0

5

UTR. In fact, 9.4 kb of the H77 genome (H product: from the very 5' end to the variable region of the 3' UTR) could be amplified from 10⁵ GE or 9.3 kb (A product: from within the 5' UTR to the variable region of the 3' UTR) from 10⁴ GE or 10⁵ GE, in a single round of long RT-PCR (Fig. 2). The PCR products amplified from 10⁵ GE of H77 were used for engineering full-length cDNA clones (see below).

10

EXAMPLE 3

Construction of Multiple Full-Length

cDNA Clones of H77 in a Single Step by

Cloning of Long RT-PCR Amplicons Directly

into a Cassette Vector with Fixed 5' and 3' Termini

15

Direct cloning of the long PCR products (H), which contained a 5' T7 promoter, the authentic 5' end, the entire ORF of H77 and a short region of the 3' UTR, into pGEM-9zf(-) vector by Not I and Xba I digestion was first 20 attempted. However, among the 70 clones examined all but two had inserts that were shorter than predicted. Sequence analysis identified a second Not I site in the majority of clones, which resulted in deletion of the nts past 25 position 9221. Only two clones (pH21, and pH50,) were missing the second Not I site and had the expected 5' and 3' sequences of the PCR product. Therefore, full-length cDNA clones (pH21 and pH50) were constructed by inserting the chosen 3' UTR into pH21, and pH50, respectively. 30 Sequence analysis revealed that clone pH21 had a complete full-length sequence of H77; this clone was tested for infectivity. The second clone, pH50, had one nt deletion

5

10

15

20

in the ORF at position 6365; this clone was used to make a cassette vector.

The complete ORF was amplified by constructing a cassette vector with fixed 5' and 3' termini as an intermediate of the full-length cDNA clones. This vector (pCV) was constructed by digestion of clone pH50 with BamHI, followed by religation, to give a shortened plasmid readily distinguished from plasmids containing the fulllength insert. Attempts to clone long RT-PCR products (H) into pCV by Age I and Afl II yielded only 1 of 23 clones with an insert of the expected size. In order to increase the efficiency of cloning, we repeated the procedure but used Pin A I and Bfr I instead of the respective isoschizomers Age I and Afl II. By this protocol, 24 of 31 H clones and 30 of 35 A clones had the full-length cDNA of H77 as evaluated by restriction enzyme digestion. A total of 16 clones, selected at random, were each retransformed, and individual plasmids were purified and completely sequenced.

EXAMPLE 4

25 <u>Demonstration of Infectious Nature</u> <u>of Transcripts of a cDNA Clone</u> <u>Representing the Consensus Sequence of Strain H77</u>

A consensus chimera was constructed from 4 of
the full-length cDNA clones with just 2 ligation steps.
The final construct, pCV-H77C, had 11 nt differences from
the consensus sequence of H77 in the ORF (Fig. 3).
However, 10 of these nucleotide differences represented
silent mutations. The chimeric clone differed from the

- 46 -

0

5

10

15

20

25

30

consensus sequence at only one amino acid [L instead of F at position 790]. Among the 18 ORFs analyzed above, the F residue was found in 11 clones and the L residue in 7 clones. However, the L residue was dominant in other isolates of genotype 1a, including a first passage of H77 in a chimpanzee (Inchauspe et al (1991)).

chimeric clone of H77 intrahepatic transfection of a chimpanzee was performed. The pCV-H77C clone was linearized with Xba I and transcribed in vitro by T7 RNA polymerase (Fig. 2). The transcription mixture was next injected into 6 sites of the liver of chimpanzee 1530. The chimpanzee became infected with HCV as measured by detection of 10² GE/ml of viral genome at week 1 p.i. Furthermore, the HCV titer increased to 10⁴ GE/ml at week 2 p.i., and reached 10⁶ GE/ml by week 8 p.i. The viremic pattern observed in the early phase of the infection with the recombinant virus was similar to that observed in chimpanzees inoculated intravenously with strain H77 or other strains of HCV (Shimizu (1990)).

The sequence of the HCV genomes from the serum sample collected at week 2 p.i. was analyzed. The consensus sequence of nts 298-9375 of the recovered genomes was determined by direct sequencing of PCR products obtained in long RT-PCR followed by nested PCR of 10 overlapping fragments. The identity to clone pCV-H77C sequence was 100%. The consensus sequence of nts 96-291,1328-1848, 3585-4106, 4763-5113 and 9322-9445 was determined from PCR products obtained in different nested RT-PCR assays. The identity of these sequences with pCV-

- 47 -

H77C was also 100%. These latter regions contained 4 mutations unique to the consensus chimera, including the artificial Afl II cleavage site in the 3' UTR. Therefore, RNA transcripts of this clone of HCV were infectious.

The infectious nature of the consensus chimera indicates that the regions of the 5' and 3' UTRs incorporated into the cassette vector do not destroy viability. This makes it highly advantageous to use the cassette vector to construct infectious cDNA clones of other HCV strains when the consensus sequence for each ORF is inserted.

In addition, two complete full-length clones (dubbed pH21 and pCV-H11) constructed were not infectious, as shown by intrahepatic injection of chimpanzees with the corresponding RNA transcripts. Thus, injection of the transcription mixture into 3 sites of the exposed liver resulted in no observable HCV replication and weekly serum samples were negative for HCV RNA at weeks 1 - 17 p.i. in a highly sensitive nested RT-PCR assay. The cDNA template injected along with the RNA transcripts was also not detected in this assay.

Moreover, the chimpanzee remained negative for antibodies to HCV throughout the follow-up. Subsequent sequence analysis revealed that 7 of 16 additional clones were defective for polyprotein synthesis and all clones had multiple amino acid mutations compared with the consensus sequence of the parent strain. For example, clone pH21, which was not infectious, had 7 amino acid substitutions in the entire predicted polyprotein compared with the consensus sequence of H77 (Fig. 3). The most

30

0

5

10

15

20

notable mutation was at position 1026, which changed L to Q, altering the cleavage site between NS2 and NS3 (Reed (1995)). Clone pCV-H11, also non-infectious, had 21 amino acid substitutions in the predicted polyprotein compared with the consensus sequence of H77 (Fig. 3). The amino acid mutation at position 564 eliminated a highly conserved C residue in the E2 protein (Okamoto (1992a)).

EXAMPLE 4A

10

15

20

25

30

5

The chimpanzee of Example 4, designated 1530, was monitored out to 32 weeks p.i. for serum enzyme levels (ALT) and the presence of anti-HCV antibodies, HCV RNA, and liver histopathology. The results are shown in Figure 18B.

A second chimp, designated 1494, was also transfected with RNA transcripts of the pCV-H77C clone and monitored out to 17 weeks p.i. for the presence of anti-HCV antibodies, HCV RNA and elevated serum enzyme levels. The results are shown in Figure 18A.

MATERIALS AND METHODS for Examples 5-10

Source Of HCV Genotype 1b

An infectious plasma pool (second chimpanzee passage) containing strain HC-J4, genotype 1b, was prepared from acute phase plasma of a chimpanzee experimentally infected with serum containing HC-J4/91 (Okamoto et al., 1992b). The HC-J4/91 sample was obtained from a first chimpanzee passage during the chronic phase of hepatitis C about 8 years after experimental infection.

PCT/US98/14688

0

10

15

20

25

The consensus sequence of the entire genome, except for the very 3' end, was determined previously for HC-J4/91 (Okamoto et al., 1992b).

5 Preparation Of HCV RNA

Viral RNA was extracted from 100 μ l aliquots of the HC-J4 plasma pool with the TRIzol system (GIBCO BRL). The RNA pellets were each resuspended in 10 μ l of 10 mM dithiothreitol (DTT) with 5% (vol/vol) RNasin (20-40 units/ μ l) (Promega) and stored at -80°C or immediately used for cDNA synthesis.

Amplification And Cloning Of The 3' UTR

A region spanning from NS5B to the conserved region of the 3' UTR was amplified in nested RT-PCR using the procedure of Yanagi et al., (1997).

minutes, and cDNA was synthesized at 42°C for 1 hour with Superscript II reverse transcriptase (GIBCO BRL) and primer H3'X58R (Table 1) in a 20 μ l reaction volume. The cDNA mixture was treated with RNase H and RNase T1 (GIBCO BRL) at 37°C for 20 minutes. The first round of PCR was performed on 2 μ l of the final cDNA mixture in a total volume of 50 μ l with the Advantage cDNA polymerase mix (Clontech) and external primers H9261F (Table 1) and H3'X58R (Table 1). In the second round of PCR [internal primers H9282F (Table 1) and H3'X45R (Table 1)], 5 μ l of the first round PCR mixture was added to 45 μ l of the PCR reaction mixture. Each round of PCR (35 cycles), was

35

- 50 -

0

5

performed in a DNA thermal cycler 480 (Perkin Elmer) and consisted of denaturation at 94°C for 1 minute (1st cycle: 1 minute 30 sec), annealing at 60°C for 1 minute and elongation at 68°C for 2 minutes. After purification with QIAquick PCR purification kit (QIAGEN), digestion with HindIII and XbaI (Promega), and phenol/chloroform extraction, the amplified products were cloned into pGEM-9zf(-) (Promega) (Yanagi et al., 1997).

10

Amplification And Cloning Of The Entire ORF

A region from within the 5' UTR to the variable region of the 3' UTR of strain HC-J4 was amplified by long RT-PCR (Fig. 1) (Yanagi et al., 1997). The cDNA was synthesized at 42°C for 1 hour in a 20 μl reaction volume with Superscript II reverse transcriptase and primer J4-9405R (5'-GCCTATTGGCCTGGAGTGGTTAGCTC-3'), and treated with RNases as above. The cDNA mixture (2 μl) was amplified by long PCR with the Advantage cDNA polymerase mix and primers A1 (Table 1) (Bukh et al., 1992; Yanagi et al., 1997) and J4-9398R (5'-

AGGATGCCTTAAGGCCTGGAGTGGTTAGCTCCCCGTTCA-3'). Primer J4-9398R contained extra bases (bold) and an artificial AflII cleavage site (underlined). A single PCR round was performed in a Robocycler thermal cycler (Stratagene), and consisted of denaturation at 99°C for 35 seconds, annealing at 67°C for 30 seconds and elongation at 68°C for 10 minutes during the first 5 cycles, 11 minutes during the next 10 cycles, 12 minutes during the following 10 cycles and 13 minutes during the last 10 cycles.

35

25

5

10

15

20

25

30

After digesting the long PCR products obtained from strain HC-J4 with PinAI (isoschizomer of AgeI) and BfrI (isoschizomer of AflII) (Boehringer Mannheim), attempts were made to clone them directly into a cassette vector (pCV), which contained the 5' and 3' termini of strain H77 (Figure 1) but no full-length clones were obtained. Accordingly, to improve the efficiency of cloning, the PCR product was further digested with BglII (Boehringer Mannheim) and the two resultant genome fragments [L fragment: PinAI/BglII, nts 156 - 8935; S fragment: BglII/BrfI, nts 8936 - 9398] were separately cloned into pCV (Figure 6).

DH5 α competent cells (GIBCO BRL) were transformed and selected on LB agar plates containing 100 μ g/ml ampicillin (SIGMA) and amplified in LB liquid cultures at 30°C for 18-20 hours.

Sequence analysis of 9 plasmids containing the S fragment (miniprep samples) and 9 plasmids containing the L fragment (maxiprep samples) were performed as described previously (Yanagi et al., 1997). Three L fragments, each encoding a distinct polypeptide, were cloned into pCV-J4S9 (which contained an S fragment encoding the consensus amino acid sequence of HC-J4) to construct three chimeric full-length HCV cDNAs (pCV-J4L2S, pCV-J4L4S and pCV-J4L6S) (Fig. 6). Large scale preparation of each clone was performed as described previously with a QIAGEN plasmid Maxi kit (Yanagi et al., 1997) and the authenticity of each clone was confirmed by sequence analysis.

PCT/US98/14688 WO 99/04008

- 52 -

Sequence Analysis

0

5

10

15

20

25

Both strands of DNA were sequenced with the ABI PRISM Dve Termination Cycle Sequencing Ready Reaction Kit using Tag DNA polymerase (Perkin Elmer) and about 90 specific sense and antisense primers. Analyses of genomic sequences, including multiple sequence alignments and tree analyses, were performed with GeneWorks (Oxford Molecular Group) (Bukh et al., 1995).

The consensus sequence of strain HC-J4 was determined by direct sequencing of PCR products (nts 11 -9412) and by sequence analysis of multiple cloned L and S The consensus sequence of the fragments (nts 156 -9371). 3' UTR (3' variable region, polypyrimidine tract and the first 16 nucleotides of the conserved region) was determined by analysis of 24 cDNA clones.

Intrahepatic Transfection Of A Chimpanzee With Transcribed RNA

Two in vitro transcription reactions were performed with each of the three full-length clones. each reaction 10 μ g of plasmid DNA linearized with Xba I (Promega) was transcribed in a 100 μ l reaction volume with T7 RNA polymerase (Promega) at 37°C for 2 hours as described previously (Yanagi et al., 1997). Five μ l of the final reaction mixture was analyzed by agarose gel 30 electrophoresis and ethidium bromide staining (Fig. 5). Each transcription mixture was diluted with 400 μ l of ice-cold phosphate-buffered saline without calcium or magnesium and then the two aliquots from the same cDNA

- 53 -

clone were combined, immediately frozen on dry ice and stored at -80°C. Within 24 hours after freezing the transcription mixtures were injected into the chimpanzee by percutaneous intrahepatic injection that was guided by ultrasound. Each inoculum was individually injected (5-6 sites) into a separate area of the liver to prevent complementation or recombination. The chimpanzee was maintained under conditions that met all requirements for its use in an approved facility.

Serum samples were collected weekly from the chimpanzee and monitored for liver enzyme levels and anti-HCV antibodies. Weekly samples of 100 μ l of serum were tested for HCV RNA in a sensitive nested RT-PCR assay (Bukh et al., 1992, Yanagi et al., 1996) with AmpliTaq Gold DNA polymerase. The genome equivalent (GE) titer of HCV was determined by testing 10-fold serial dilutions of the extracted RNA in the RT-PCR assay (Yanagi et al., 1996) with 1 GE defined as the number of HCV genomes present in the highest dilution which was positive in the RT-nested PCR assay.

infectious in vivo, the NS3 region (nts 3659 - 4110) from the chimpanzee serum was amplified in a highly sensitive and specific nested RT-PCR assay with AmpliTaq Gold DNA polymerase and the PCR products were cloned with a TA cloning kit (Invitrogen). In addition, the consensus sequence of the nearly complete genome (nts 11 - 9441) was determined by direct sequencing of overlapping PCR products.

0

5

10

15

- 54 -

0

EXAMPLE 5

Sequence Analysis Of Infectious Plasma Pool Of Strain HC-J4 Used As The Cloning Source

5

10

15

As an infectious cDNA clone of a genotype la strain of HCV had been obtained only after the ORF was engineered to encode the consensus polypeptide (Kolykhalov et al., 1997; Yanagi et al., 1997), a detailed sequence analysis of the cloning source was performed to determine the consensus sequence prior to constructing an infectious cDNA clone of a 1b genotype.

A plasma pool of strain HC-J4 was prepared from a cute phase plasmapheresis units collected from a chimpanzee experimentally infected with HC-J4/91 (Okamoto et al., 1992b). This HCV pool had a PCR titer of 10^4 - 10^5 GE/ml and an infectivity titer of approximately 10^3 chimpanzee infectious doses per ml.

20

25

The heterogeneity of the 3' UTR of strain HC-J4 was determined by analyzing 24 clones of nested RT-PCR product. The consensus sequence was identical to that previously published for HC-J4/91 (Okamoto et al., 1992b), except at position 9407 (see below). The variable region consisted of 41 nucleotides (nts. 9372 - 9412), including two in-frame termination codons. Furthermore, its sequence was highly conserved except at positions 9399 (19 A and 5 T clones) and 9407 (17 T and 7 A clones). The poly U-UC region varied slightly in composition and greatly in length (31-162 nucleotides). In the conserved region, the first 16 nucleotides of 22 clones were identical to those previously published for other genotype

35

- 55 -

1 strains, whereas two clones each had a single point mutation. These data suggested that the structural organization at the 3' end of HC-J4 was similar to that of the infectious clone of a genotype la strain of Yanagi et al (1997).

Next, the entire ORF of HC-J4 was amplified in a single round of long RT-PCR (Figure 5). The original plan was to clone the resulting PCR products into the *PinAI* and *BrfI* site of a HCV cassette vector (pCV), which had fixed 5' and 3' termini of genotype 1a (Yanagi et al., 1997) but since full-length clones were not obtained, two genome fragments (L and S) derived from the long RT-PCR products (Figure 6) were separately subcloned into pCV.

To determine the consensus sequence of the ORF, the sequence of 9 clones each of the L fragment (pCV-J4L) and of the S fragment (pCV-J4S) was determined and quasispecies were found at 275 nucleotide (3.05 %) and 78 amino acid (2.59 %) positions, scattered throughout the 9030 nts (3010 aa) of the ORF (Figure 7). Of the 161 nucleotide substitutions unique to a single clone, 71% were at the third position of the codon and 72 % were silent.

Each of the nine L clones represented the near complete ORF of an individual genome. The differences among the L clones were 0.30 - 1.53% at the nucleotide and 0.31 - 1.47% at the amino acid level (Figure 8). Two clones, L1 and L7, had a defective ORF due to a single nucleotide deletion and a single nucleotide insertion, respectively. Even though the HC-J4 plasma pool was obtained in the early acute phase, it appeared to contain

30

5

10

15

20

•

5

at least three viral species (Figure 9). Species A contained the L1, L2, L6, L8 and L9 clones, species B the L3, L7 and L10 clones and species C the L4 clone. Although each species A clone was unique all A clones differed from all B clones at the same 20 amino acid sites and at these positions, species C had the species A sequence at 14 positions and the species B sequence at 6 positions (Figure 7).

Okamoto and coworkers (Okamoto et al., 1992b)

previously determined the nearly complete genome consensus sequence of strain HC-J4 in acute phase serum of the first chimpanzee passage (HC-J4/83) as well as in chronic phase serum collected 8.2 years later (HC-J4/91). In addition, they determined the sequence of amino acids 379 to 413 (including HVR1) and amino acids 468 to 486 (including HVR2) of multiple individual clones (Okamoto et al., 1992b).

It was found by the present inventors that the sequences of individual genomes in the plasma pool collected from a chimpanzee inoculated with HC-J4/91 were all more closely related to HC-J4/91 than to HC-J4/83 (Figures 8, 9) and contained HVR amino acid sequences closely related to three of the four viral species previously found in HC-J4/91 (Figure 10).

Thus, the data presented herein demonstrate the occurrence of the simultaneous transmission of multiple species to a single chimpanzee and clearly illustrates the difficulties in accurately determining the evolution of HCV over time since multiple species with significant changes throughout the HCV genome can be present from the

30

20

5

15

20

25

- 57 -

onset of the infection. Accordingly, infection of chimpanzees with monoclonal viruses derived from the infectious clones described herein will make it possible to perform more detailed studies of the evolution of HCV in vivo and its importance for viral persistence and pathogenesis.

EXAMPLE 6

10 Determination Of The Consensus Sequence Of HC-J4 In The Plasma Pool

> The consensus sequence of nucleotides 156-9371 of HC-J4 was determined by two approaches. approach, the consensus sequence was deduced from 9 clones of the long RT-PCR product. In the other approach the long RT-PCR product was reamplified by PCR as overlapping fragments which were sequenced directly. The two "consensus" sequences differed at 31 (0.34%) of 9216 nucleotide positions and at 11 (0.37%) of 3010 deduced amino acid positions (Figure 7). At all of these positions a major quasispecies of strain HC-J4 was found in the plasma pool. At 9 additional amino acid positions the cloned sequences displayed heterogeneity and the direct sequence was ambiguous (Figure 7). Finally, it should be noted that there were multiple amino acid positions at which the consensus sequence obtained by direct sequencing was identical to that obtained by cloning and sequencing even though a major quasispecies was detected (Figure 7).

For positions at which the two "consensus" sequences of HC-J4 differed, both amino acids were

- 58 -

0

5

included in a composite consensus sequence (Figure 7).

However, even with this allowance, none of the 9 L clones analyzed (aa 1 - 2864) had the composite consensus sequence: two clones did not encode the complete polypeptide and the remaining 7 clones differed from the consensus sequence by 3 - 13 amino acids (Figure 7).

EXAMPLE 7

10

Construction Of Chimeric Full-Length cDNA Clones Containing The Entire ORF Of HC-J4

The cassette vector used to clone strain H77 was used to construct an infectious cDNA clone containing the ORF of a second subtype.

15

20

25

30

In brief, three full-length cDNA clones were constructed by cloning different L fragments into the PinAI/BglII site of pCV-J4S9, the cassette vector for genotype 1a (Figure 6), which also contained an S fragment encoding the consensus amino acid sequence of HC-J4. Therefore, although the ORF was from strain HC-J4, most of the 5' and 3' terminal sequences originated from strain H77. As a result, the 5' and 3' UTR were chimeras of genotypes 1a and 1b (Figure 11).

The first 155 nucleotides of the 5' UTR were from strain H77 (genotype 1a), and differed from the authentic sequence of HC-J4 (genotype 1b) at nucleotides 11, 12, 13, 34 and 35. In two clones (pCV-J4L2S, pCV-J4L6S) the rest of the 5' UTR had the consensus sequence of HC-J4, whereas the third clone (pCV-J4L4S) had a single nucleotide insertion at position 207. In all 3 clones the first 27 nucleotides of the 3' variable region of the 3'

- 59 -

UTR were identical with the consensus sequence of HC-J4. The remaining 15 nucleotides of the variable region, the poly U-UC region and the 3' conserved region of the 3' UTR had the same sequence as an infectious clone of strain H77 (Figure 11).

None of the three full-length clones of HC-J4 had the ORF composite consensus sequence (Figures 7, 12). The pCV-J4L6S clone had only three amino acid changes: Q for R at position 231 (E1), V for A at position 937 (NS2) and T for S at position 1215 (NS3). The pCV-J4L4S clone had 7 amino acid changes, including a change at position 450 (E2) that eliminated a highly conserved N-linked glycosylation site (Okamoto et al., 1992a). Finally, the pCV-J4L2S clone had 9 amino acid changes compared with the consensus sequence of HC-J4. A change at position 304 (E1) mutated a highly conserved cysteine residue (Bukh et al., 1993; Okamoto et al., 1992a).

20

5

10

15

EXAMPLE 8

Transfection Of A Chimpanzee By In Vitro Transcripts Of A Chimeric cDNA

was determined by ultra-sound-guided percutaneous intrahepatic injection into the liver of a chimpanzee of the same amount of cDNA and transcription mixture for each of the clones (Figure 5). This procedure is a less invasive procedure than the laparotomy procedure utilized by Kolykhalov et al. (1997) and Yanagi et al. (1997) and should facilitate in vivo studies of cDNA clones of HCV in

- 60 -

chimpanzees since percutaneous procedures, unlike laparotomy, can be performed repeatedly.

As shown in Figure 13, the chimpanzee became infected with HCV as measured by increasing titers of 10^2 GE/ml at week 1 p.i., 10^3 GE/ml at week 2 p.i. and 10^4 - 10^5 GE/ml during weeks 3 to 10 p.i.

The viremic pattern found in the early phase of the infection was similar to that observed for the recombinant H77 virus in chimpanzees (Bukh et al., unpublished data; Kolykhalov et al., 1997; Yanagi et al., 1997). The chimpanzee transfected in the present study was chronically infected with hepatitis G virus (HGV/GBV-C) (Bukh et al., 1998) and had a titer of 106 GE/ml at the time of HCV transfection. Although HGV/GBV-C was originally believed to be a hepatitis virus, it does not cause hepatitis in chimpanzees (Bukh et al., 1998) and may not replicate in the liver (Laskus et al., 1997). The present study demonstrated that an ongoing infection of HGV/GBV-C did not prevent acute HCV infection in the chimpanzee model.

However, to identify which of the three fulllength HC-J4 clones were infectious, the NS3 region (nts.
3659 - 4110) of HCV genomes amplified by RT-PCR from serum
samples taken from the infected chimpanzee during weeks 2
and 4 post-infection (p.i.) were cloned and sequenced. As
the PCR primers were a complete match with each of the
original three clones, this assay should not have
preferentially amplified one virus over another. Sequence
analysis of 26 and 24 clones obtained at weeks 2 and 4

5

10

15

20

25

- 61 -

p.i., respectively, demonstrated that all originated from the transcripts of pCV-J4L6S.

Moreover, the consensus sequence of PCR products of the nearly complete genome (nts. 11-9441), amplified from serum obtained during week 2 p.i., was identical to the sequence of pCV-J4L6S and there was no evidence of quasispecies. Thus, RNA transcripts of pCV-J4L6S, but not of pCV-J4L2S or pCV-J4L4S, were infectious in vivo. The data in Figure 13 is therefore the product of the transfection of RNA transcripts of pCV-J4L6S.

In addition, the chimeric sequences of genotypes la and 1b in the UTRs were maintained in the infected chimpanzee. The consensus sequence of nucleotides 11 -15 341 of the 5' UTR and the variable region of the 3' UTR, amplified from serum obtained during weeks 2 and 4 p.i., had the expected chimeric sequence of genotypes la and lb (Fig. 11). Also three of four clones of the 3' UTR obtained at week 2 p.i. had the chimeric sequence of the 20 variable region, whereas a single substitution was noted in the fourth clone. However, in all four clones the poly U region was longer (2-12 nts) than expected. Also, extra C and G residues were observed in this region. For the 25 most part, the number of C residues in the poly UC region was maintained in all clones although the spacing varied. As shown previously, variations in the number of U residues can reflect artifacts introduced during PCR 30 amplification (Yanagi et al., 1997). The sequence of the first 19 nucleotides of the conserved region was maintained in all four clones. Thus, with the exception of the poly U-UC region, the genomic sequences recovered

0

5

- 62 -

0

5

from the infected chimpanzee were exactly those of the chimeric infectious clone pCV-J4BL6S.

The results presented in Figure 13 therefore demonstrate that HCV polypeptide sequences other than the consensus sequence can be infectious and that a chimeric genome containing portions of the H77 termini could produce an infectious virus. In addition, these results showed for the first time that it is possible to make infectious viruses containing 5' and 3' terminal sequences specific for two different subtypes of the same major genotype of HCV.

EXAMPLE 9

15

20

10

Construction Of A Chimeric 1a/lb Infectious Clone

A chimeric 1a/1b infectious clone in which the structural region of the genotype 1b infectious clone is inserted into the 1a clone of Yanagi et al. (1997) is constructed by following the protocol shown in Figure 15. The resultant chimera contains nucleotides 156-2763 of the 1b clone described herein inserted into the 1a clone of Figures 4A-4F. The sequences of the primers shown in Figure 15 which are used in constructing this chimeric clone, designated pH77CV-J4, are presented below.

30

25

- 63 -

0

- 1. H2751S (Cla I/Nde I)
 CGT CAT CGA TCC TCA GCG GGC ATA TGC ACT GGA CAC GGA
- 2. <u>H2870R</u>
 CAT GCA CCA GCT GAT ATA GCG CTT GTA ATA TG
- 5 3. <u>H7851S</u> TCC GTA GAG GAA GCT TGC AGC CTG ACG CCC
 - 4. <u>H9173 R(P-M)</u>
 GTA CTT GCC ACA TAT AGC AGC CCT GCC TCT G
- 10 5. H9140S (P-M)
 CAG AGG AGG CAG GGC TGC TAT ATG TGG CAA GTA C
 - 6. <u>H9417R</u>
 CGT CTC TAG ACA GGA AAT GGC TTA AGA GGC CGG AGT GTT
 TAC C
- 7. $\frac{\text{J4-2271S}}{\text{TGC AAT TGG ACT CGA GGA GAG CGC TGT AAC TTG GAG}}$
 - 8. <u>J4-2776R (Nde I)</u> CGG TCC AAG GCA TAT GCT CGT GGT AAC GCC AG

Transcripts of the chimeric la/lb clone (whose

sequence is shown in Figures 16A-16F) are then produced
and transfected into chimpanzees by the methods described
in the Materials and Methods section herein and the
transfected animals are then be subjected to biochemical

(ALT levels), histopathological and PCR analyses to
determine the infectivity of the chimeric clone.

EXAMPLE 10

30 Construction of 3' Deletion Mutants Of The la Infectious Clone pCV-H77C

Seven constructs having various deletions in the 3' untranslated region (UTR) of the 1a infectious clone pCV-H77C were constructed as described in Figures 17A-17B.

- 64 -

0

The 3' untranslated sequence remaining in each of the seven constructs following their respective deletions is shown in Figures 17A-17B.

5

10

15

Construct pCV-H77C(-98X) containing a deletion of the 3'-most 98 nucleotide sequences in the 3'-UTR was transcribed in vitro according to the methods described herein and 1 ml of the diluted transcription mixture was percutaneously transfected into the liver of a chimpanzee with the aid of ultrasound. After three weeks, the transfection was repeated. The chimpanzee was observed to be negative for hepatitis C virus replication as measured by RT-PCR assay for 5 weeks after transfection. These results demonstrate that the deleted 98 nucleotide 3'-UTR sequence was critical for production of infectious HCV and appear to contradict the reports of Dash et al. (1996) and Yoo et al. (1995) who reported that RNA transcripts from cDNA clones of HCV-1 and HCV-N lacking the terminal 98 conserved nucleotides at the very 3' end of the 3'-UTR

20

25

Transcripts of the (-42X) mutant (Figure 17C) were also produced and transfected into a chimpanzee and transcripts of the other five deletion mutants shown in Figures 17D-17G) are to be produced and transfected into chimpanzees by the methods described herein. All transfected animals are to then be assayed for viral replication via RT-PCR.

resulted in viral replication after transfection into

human hematoma cell lines.

- 65 -

0

5

10

15

20

25

Discussion

In two recent reports on transfection of chimpanzees, only those clones engineered to have the independently determined and slightly different consensus amino acid sequence of the polypeptide of strain H77 were infectious (Kolykhalov et al., 1997; Yanagi et al., 1997). Although the two infectious clones differed at four amino acid positions, these differences were represented in a major component of the quasispecies of the cloning source. In the present study, a single consensus sequence of strain HC-J4 could not be defined because the consensus sequence obtained by two different approaches (direct sequencing and sequencing of cloned products) differed at 20 amino acid positions, even though the same genomic PCR product was analyzed. The infectious clone differed at two positions from the composite amino acid consensus sequence, from the sequence of the 8 additional HC-J4 clones analyzed in this study and from published sequences of earlier passage samples. An additional amino acid differed from the composite consensus sequence but was found in two other HC-J4 clones analyzed in this study. The two non-infectious full-length clones of HC-J4 differed from the composite consensus sequence by only 7 and 9 amino acid differences. However, since these clones had the same termini as the infectious clone (except for a single nucleotide insertion in the 5' UTR of pCV-J4L4S), one or more of these amino acid changes in each clone was apparently deleterious for the virus.

35

- 66 -

It was also found in the present study that HC-J4, like other strains of genotype 1b (Kolykhalov et al., 1996; Tanaka et al., 1996; Yamada et al., 1996), had a poly U-UC region followed by a terminal conserved element. The poly U-UC region appears to vary considerably so it was not clear whether changes in this region would have a significant effect on virus replication. On the other hand, the 3'98 nucleotides of the HCV genome were previously shown to be identical among other strains of genotypes 1a and 1b (Kolykhalov et al., 1996; Tanaka et al., 1996). Thus, use of the cassette vector would not alter this region except for addition of 3 nucleotides found in strain H77 between the poly UC region and the 3'98 conserved nucleotides.

In conclusion, an infectious clone representing a genotype 1b strain of HCV has been constructed. Thus, it has been demonstrated that it was possible to obtain an infectious clone of a second strain of HCV. In addition, it has been shown that a consensus amino acid sequence was not absolutely required for infectivity and that chimeras between the UTRs of two different genotypes could be viable.

5

REFERENCES

Alter, M. J. (1997). <u>Hepatology</u> **26**, 62S-65S. Blight, K. J. and Rice, C. M. (1997). <u>J. Virol.</u> **71**, 7345-7352.

Brechot, C. (1997). Hepatology 25, 772-774.

Bukh, J., et al. (1992). <u>Proc. Natl. Acad. Sci. USA</u> 89, 187-191.

Bukh, J., et al. (1993). <u>Proc. Natl. Acad. Sci. USA</u>

90, 8234-8238.

Bukh, J., et al. (1995). <u>Semin. Liver Dis.</u> **15**, 41-63.

Bukh, J., et al. (1997). Genetic heterogeneity of
hepatitis C virus and related viruses. In "Viral
Hepatitis and Liver Disease, Proceedings of IX Triennial
International Symposium on Viral Hepatitis and Liver
Disease, Rome, Italy, 1996" (M. Rizzetto, R. H. Purcell,
J. L. Gerin and G. Verme, Eds.), pp. 167-175. Edizioni
Minerva Medica, Turin.

Bukh, J., et al. (1998). Experimental infection of chimpanzees with hepatitis G virus and genetic analysis of the virus. J. Infect. Dis. (in press).

Choo, Q.-L., et al. (1991). Diversity of the hepatitis C virus. Proc. Natl. Acad. Sci. USA 88, 2451-2455.

Dash, S., et al. (1997). <u>Am. J. Pathol.</u> **151,** 363-373.

Emerson, S.U. et al (1992) <u>J. Virol.</u>, 66:6649-6654.

Farci, P. and Purcell, R. H. (1993). Natural history and experimental models. In "Viral Hepatitis: Scientific Basis and Clinical Management". (A. J. Zuckerman and H.

35

5

C. Thomas, Eds.). pp. 241-267. Churchill Livingstone, Edinburgh.

Farci, P., et al. (1994). Prevention of hepatitis C virus infection in chimpanzees after antibody-mediated <u>in vitro</u> neutralization. <u>Proc. Natl. Acad. Sci. USA</u> 91, 7792-7796.

Farci, P., et al. (1996). <u>Proc. Natl. Acad. Sci. USA</u> 93, 15394-15399.

Farci, P., et al. (1997). <u>Springer Semin.</u>

Immunopathol. 19, 5-26.

Fausto, N. (1997). Am. J. Pathol. 151, 361.

Feinstone, S.M. et al (1981) <u>J. Infect. Dis.</u> 144:588-598.

15 Fried, M. W. and Hoofnagle, J. H. (1995). <u>Semin.</u> Liver Dis. **15**, 82-91.

Han, J.M. et al (1992) Nuc. Acids Res., 20:3250.

Hijikata, M., et al. (1991). <u>Biochem. Biophys. Res.</u>

20 <u>Commun.</u> 175, 220-228.

Honda, M., et al. (1996). RNA 2, 955-968.

Hoofnagle, J. H. (1997). Hepatology 26, 15S-20S.

Houghton, M. (1996). Hepatitis C viruses. In

"Fields Virology" (B. N. Fields, D. M. Knipe, P. M. Howley, et al., Eds.), Third ed., Lippincott-Raven

Publishers, Philadelphia.

Inchauspe, G. et al. (1991) <u>Proc. Natl. Acad. Sci.</u>
U.S.A., 88:10292-10296.

30 Ito, T. and Lai, M. M. C. (1997). <u>J. Virol.</u> **71**, 8698-8706.

Kato, N., et al. (1990). Proc. Natl. Acad. Sci. USA
87, 9524-9528.

0

Kolykhalov, A. A., Feinstone, S. M. and Rice, C. M. (1996). J. Virol. 70, 3363-3371.

Kolykhalov, A. A., et al. (1997). <u>Science</u> **277**, 570-574.

Krieg, A.M. et al. (1995) <u>Nature</u>, 374:546.
Krieg, A.M. et al. (1996) <u>J. Lab. Clin. Med.</u>,
128:128.

Laskus, T., et al. (1997). <u>J. Virol.</u> **71,** 7804-7806. Major, M.E. and Feinstone, S.M. (1997) <u>Heptology</u>

10 Major, M.E. and Feinstone, S.M. (1997) <u>Heptology</u> 25:1527-1538.

Ogata, N. et al (1991) <u>Proc. Natl. Acad. Sci. U.S.A.</u>, 88:3392-3396.

Okamoto, H., et al. (1992a). <u>Virology</u> 188, 331-341.

Okamoto, H., et al. (1992b) <u>Virology</u> 190, 894-899.

Reed, K.E., et al. (1995) <u>J. Virol.</u>, 69:4127-4136.

Rice, C. M. (1996). Flaviviridae: The viruses and their replication, In "Fields Virology". (B. N. Fields,

D. M. Knipe, P. M. Howley, et al., Eds.), Third ed., Lippincott-Raven Publishers, Philadelphia.

Shimizu, Y.K., et al. (1990) <u>Proc. Natl. Acad. Sci.</u> <u>U.S.A.</u>, **87**:6441-6444.

Shimizu, Y. K., et al. (1992). <u>Proc. Natl. Acad.</u> <u>Sci. USA</u> **89**, 5477-5481.

Shimizu, Y. K., et al. (1996). <u>Virology</u> 223, 409-412.

Simmonds, P., et al. (1993). <u>J. Gen. Virol.</u> **74,** 2391-2399.

Takamizawa, A., et al. (1991). <u>J. Virol.</u> **65,** 1105-1113.

- 70 -

0

5

Tanaka, T., et al. (1995). <u>Biochem. Biophys. Res.</u>
Commun. **215**, 744-749.

Tanaka, T., et al. (1996). Structure of the 3' terminus of the hepatitis C virus genome. <u>J. Virol.</u> 70, 3307-3312.

Tellier, R. et al (1996) <u>Proc. Natl. Acad. Sic.</u> <u>U.S.A.</u>, **93**:4370-4373.

Tellier, R., et al (1996a) J. Clin. Microbiol,

34:3085-3091.

Tsuchihara, K., et al. <u>J. Virol.</u> **71,** 6720-6726.

Tsukiyama-Kohara, K., et al. <u>J. Virol.</u> **66,** 1476-

Weiner, A. J., et al. (1991). <u>Virology</u> **180,** 842-848.

Weiner, A. J., et al. (1992). <u>Proc. Natl. Acad. Sci.</u>

USA **89,** 3468-3472.

Yamada, N., et al. (1996). <u>Virology</u> 223, 255-261. Yanagi, M., et al. (1996). <u>J. Infect. Dis.</u> 174,

20 1324-1327.

Yanagi, M., et al. (1997). <u>Proc. Natl. Acad. Sci.</u> <u>USA</u> **94**, 8738-8743.

Yoo, B. J., et al. (1995). <u>J. Virol.</u> **69**, 32-38.

25

30

- 71 -

5

10

20

25

WHAT IS CLAIMED IS:

1. A purified and isolated nucleic acid molecule which encodes human hepatitis C virus, said molecule capable of expressing said virus when transfected into cells.

- 2. The nucleic acid molecule of claim 1, wherein said molecule encodes the amino acid sequence shown in Figures 14G-14H.
- 3. The nucleic acid molecule of claim 2, wherein said molecule comprises the nucleic acid sequence shown in Figures 14A-14F.
- 4. The nucleic acid molecule acid molecule of claim 1, wherein said molecule encodes the amino acid sequence shown in Figures 4G-4H.
 - 5. The nucleic acid molecule of claim 4, wherein said molecule comprises the nucleic acid sequence shown in Figures 4A-4F.
 - 6. The nucleic acid molecule of claim 1, wherein a fragment of said molecule which encodes the structural region of hepatitis C virus has been replaced by the structural region from the genome of another hepatitis C virus strain.
 - 7. The nucleic acid molecule of claim 6, wherein said molecule encodes the amino acid sequence shown in Figures 16G-16H.
- 8. The nucleic acid molecule of claim 7, wherein said molecule comprises the nucleic acid sequence shown in Figures 16A-16F.

- 72 -

0

5

15

- 9. The nucleic acid molecule of claim 1, wherein a fragment of the nucleic acid molecule which encodes at least one HCV protein has been replaced by a fragment of the genome of another hepatitis C virus strain which encodes the corresponding protein.
- 10. The nucleic acid molecule of claim 9, wherein the protein is selected from the group consisting of E1, E2 or NS4 proteins.
- 11. The nucleic acid molecule of claim 1, wherein a fragment of the molecule encoding all or part of an HCV protein has been deleted.
 - 12. The nucleic acid molecule of claim 11, wherein the HCV protein is selected from the group consisting of P7, NS4B or NS5A proteins.
 - 13. A DNA construct comprising a nucleic acid molecule according to claims 1, 3, 5 or 8.
- 14. An RNA transcript of the DNA construct of
 20 claim 13.
 - 15. A cell transfected with the DNA construct of claim 13.
 - 16. A cell transfected with RNA transcript of claim 14.
 - 17. A hepatitis C virus polypeptide produced by the cell of claim 15.
 - 18. A hepatitis C virus polypeptide produced by the cell of claim 16.
- 30 19. A hepatitis C virus produced by the cell of claim 13.
 - 20. A hepatitis C virus produced by the cell of claim 14.

0

10

15

20

25

21. A hepatitis C virus whose genome comprises a nucleic acid molecule according to claims 1, 3, 5, 6, 8, or 9.

- 5 comprising transfecting a host cell with the RNA transcript of claim 14.
 - 23. A polypeptide encoded by a nucleic acid sequence according to claims 1, 2, 4 or 7 or a fragment thereof.
 - 24. The polypeptide of claim 23, wherein said polypeptide is selected from the group consisting of NS3 protease, El protein, E2 protein or NS4 protein.
 - 25. A method for assaying candidate antiviral agents for activity against HCV, comprising
 - a) exposing a cell containing the hepatitis C virus of claim 21 to the candidate antiviral agent; and
 - b) measuring the presence or absence of hepatitis C virus replication in the cell of step (a).
 - 26. The method of claim 25, wherein said replication in step (b) is measured by at least one of the following: negative strand RT-PCR, quantitative RT-PCR, Western blot, immunofluoresence, or infectivity in a susceptible animal.
 - 27. A method for assaying candidate antiviral agents for activity against HCV, comprising:
 - a) exposing an HCV protease encoded by a nucleic acid sequence according to claims 1, 2, 4, or 7, or a fragment thereof to the

WO 93/04000

- 74 -

0	

candidate antiviral agent in the presence of a protease substrate; and

b) measuring the protease activity of said protease.

5

28. The method of claim 27, wherein said HCV protease is selected from the group consisting of an NS3 domain protease, an NS3-NS4A fusion polypeptide, or an NS2-NS3 protease.

10

- 29. An antiviral agent identified as having antiviral activity for HCV by the method of claim 25.
- 30. An antiviral agent identified as having antiviral activity for HCV by the method of claim 27.
 - 31. Antibody to the polypeptide of claim 23.

15

20

25

21.

- 32. Antibody to the hepatitis C virus of claim
- 33. A method for determining the susceptibility of cells in vitro to support HCV infection, comprising the steps of:
 - a. growing animal cells in vitro;

b. transfecting into said cells the nucleic acid of claim 1; and

c. determining if said cells show indicia of HCV replication.

34. The method according to claim 33, wherein said cells are human cells.

35. A cassette vector for cloning viral genomes, comprising, inserted therein, the nucleic acid sequence according to claim 2, said vector reading in the

0

5

15

20

25

correct phase for the expression of said inserted sequence and having an active promoter sequence upstream thereof.

- 36. The cassette vector of claim 35, wherein the cassette vector is produced from plasmid pCV.
- 37. The cassette vector of claim 35, wherein the vector also contains one or more expressible marker genes.
- 38. The cassette vector of claim 35, wherein the inserted DNA sequence contains at least one ORF of the HCV genome from any strain.
 - 39. The cassette vector of claim 35, wherein the promoter is a bacterial promoter.
 - 40. A composition comprising a polypeptide of claim 23 suspended in a suitable amount of a pharmaceutically acceptable diluent or excipient.
 - 41. A method for treating hepatitis C viral infection comprising the administration to a animal in need thereof of a clinically effective amount of the composition of claim 40.
 - 42. A composition comprising a nucleic acid molecule of claim 1 suspended in a suitable amount of a pharmaceutically acceptable diluent or excipient.
 - 43. A method for treating hepatitis C viral infection comprising the administration to an animal in need thereof of a clinically effective amount of the composition of claim 42.

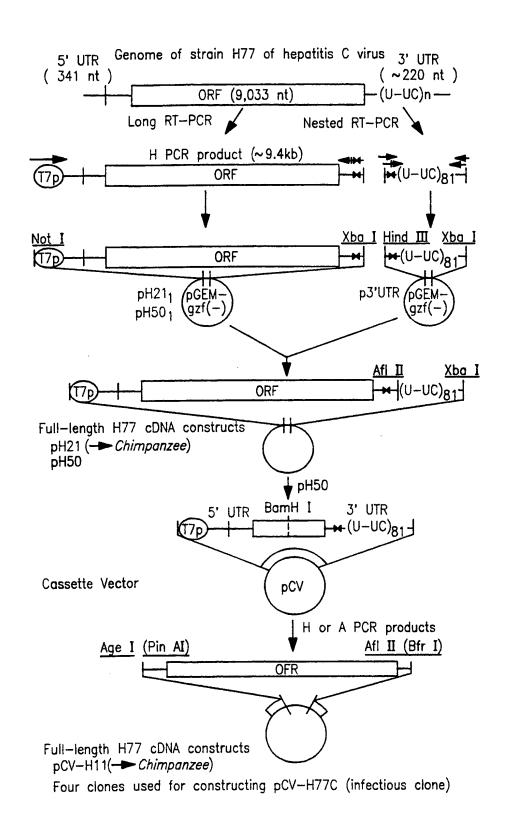
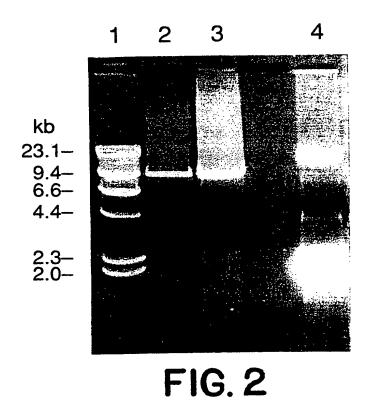
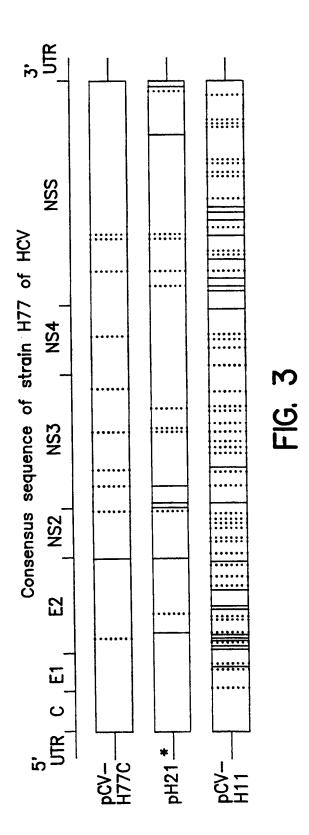


FIG. I



SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)

H77C

10 20 30 40 50 1234567890 1234567890 1234567890 1234567890 GCCAGCCCC TGATGGGGGC GACACTCCAC CATGAATCAC TCCCCTGTGA 50
GCCAGCCCC TGATGGGGC GACACTCCAC CATGAATCAC TCCCCTGTGA 50
GGAACTACIG TCTTCACGCA GAAAGCGTCT AGCCATGGCG TTAGTATGAG 100
TGICGIGCAG CCICCAGGAC CCCCCCTCCC GGGAGAGCCA TAGIGGICIG 150
CCCAACCECT CAGTACACCE CAATTCCCAG CACCACCEC TCCTTTCTTG 200
GATAAACCCG CICAATGCCT GCAGATTIGG GCGIGCCCCC GCAAGACIGC 250
TAGCCGAGIA GIGITGGGIC GCGAAAGGCC TIGIGGIACT GCCIGATAGG 300
GIGCTIGGGA GIGCCCCGGG AGGICTCGTA GACCGIGCAC CATGAGCACG 350
AATCCTAAAC CTCAAAGAAA AACCAAACGT AACACCAACC GTCGCCCACA 400
GGACGICAAG TICCCGGGIG GCGGICAGAT CGITGGIGGA GITTACTIGT 450
TGCCGCGCAG GGGCCCTAGA TTGGGTGTGC GCGCGACGAG GAAGACTTCC 500
CACCOGICCC AACCICCACG TAGACGICAG CCIATCCCCA ACCCACGICG 550
GCCCGAGGCC AGGACCIGGG CICAGCCCGG GIACCCITGG CCCCICIATG 600
GCAATGAGGG TIGCGGGTGG GCGGGATGGC TCCTGTCTCC CCGTGGCTCT 650
CGCCTAGCT GGGGCCCCAC AGACCCCCGG CGTAGGTCGC GCAATTIGGG 700
TAAGGICATC GATACCCTTA CGIGCGGCTT CGCCGACCTC ATGGGGTACA 750
TACCECTOGT CECCCCCCT CTTCCACCCC CTCCCACCCC CCTCCCCAT 800
GCCGICCGCG TICIGGAAGA CGCCGIGAAC TATGCAACAG GGAACCITCC 850
TEGTTECTOT TTCTCTATCT TCCTTCTGGC CCTGCTCTCT TCCCTGACTG 900
TGCCCGCTTC AGCCTACCAA GIGCGCAATT CCICGGGGCT TTACCATGIC 950
ACCAATGATT GCCCTAACTC GAGTATTGTG TACGAGGGGG CCGATGCCAT 1000
CCTGCACACT CCGGGGTGTG TCCCTTGCGT TCGCGAGGGT AACGCCTCGA 1050
CONCRETE CONTROL CONTR
CCCACAACGC AGCTTCGACG TCATATCGAT CTGCTTGTCG GGAGCGCCAC 1150
CCICIGCICG CCCCICIACG TGGGGGACCT GIGGGGGTCT GICITICITG 1200
TIGGICAACT GITTACCTIC TCTCCCAGGC GCCACTGGAC GACGCAAGAC 1250
TGCAATTGTT CTATCTATCC CGGCCATATA ACGGTCATC GCATGGCATG
GCATATGATG ATGAACTGGT CCCCTACGGC AGCGTTGGTG GTAGCTCAGC 1350
TOCTOCOGENT COCACAAGCC ATCATOGEACA TGATOGCTGG TGCTCACTGG 1400
GGAGICCIGG CGGCATAGC GIATTICICC AIGGIGGGGA ACIGGGCGAA 1450
GCICCIGGIA GIGCIGCICC TATTIGCCG CGICGACGC GAAACCCACG 1500
TCACCGGGG AAATGCCGGC CGCACCACGG CTGGGCTTGT TGGTCTCCTT 1550
ACACCAGGG CCAAGCAGAA CATCCAACTG ATCAACACCA ACGGCAGTTG 1600
GCACATCAAT AGCACGGCCT TGAATTGCAA TGAAAGCCTT AACACCGGCT 1650
GGITAGCAGG GCICITCIAT CAACACAAAT TCAACICITC AGGCIGICCT 1700
CACACCTICG CCACCTGCCG ACCCCTTACC CATTTTGCCC ACCCCTGCG 1750
TOCTATOAGT TATGOCAACG GAAGCGGCCT CGACGAACGC CCCTACTGCT 1800
GCACIACCC TCCAAGACCT TGIGGCATIG TGCCCGCAAA GAGCGIGIGI 1850
GCCCGGIAT ATTGCTTCAC TCCCAGCCCC GTGGTGGTGG GAACGACCGA 1900

FIG. 4A

10	20	30	40	FA	
1234567890	20 1234567890		40 1234567890	50 1234567890	
			TGCAAATGAT		1950
			GCAATIGGIT		2000
			TGCGGAGCGC		2050
			CIGCCCCACT		2100
GCAAACATCC	GGAAGCCACA	TACTCTCGGT	GCCCTCCCC	TCCCTGGATT	2150
ACACCCAGGT	GCATGGTCGA	CIACCOGIAT	AGGCTTTGGC	ACTATOCTEG	2200
TACCATCAAT	TACACCATAT	TCAAAGICAG	GATGIACGIG	GCAGGGTOG	2250
AGCACAGGCT	GGAAGCGGCC	TGCAACTGGA	CGCGGGGGGA	ACCCIGICAT	2300
CIGGAAGACA	GGGACAGGIC	CGAGCTCAGC	CCGITCCICC	TGTCCACCAC	2350
ACAGIGGCAG	GICCITCCGT	GITCITTCAC	GACCCIGCCA	GCCTTGTCCA	2400
CCCCCCTCAT	CCACCTCCAC	CAGAACATTG	TGGACGIGCA	GIACITGIAC	2450
GGGGTAGGGT	CAAGCATCGC	GICCIGGGCC	ATTAAGTGGG	AGTACGTCGT	2500
TCTCCTGTTC	CITCICCITG	CAGAGGGGG	CETCTCCTCC	TGCTTGTGGA	2550
TGATGITACI	CATATCCCAA	GCGGAGGCGG	CITTGGAGAA	CCTCGTAATA	2600
CICAAIGCAG	CATCCCTGGC	CGGGACGCAC	GCTTGTGT	CCTTCCTCGT	2650
GITCITCIGC	TTTGCGTGGT	ATCIGAAGGG	TAGGIGGGIG	CCCCCCACCCCC	2700
TCTACGCCCT	CTACGGGATG	TEGECTICICC	TCCTGCTCCT	CCICCCCTIC	2750
CCTCAGCGGG	CATACGCACT	GGACACGGAG	GIGGCCGCGI	CCICICCCCC	2800
CGITGITCIT	GICGGGITAA	TGGCGCTGAC	TCIGICGCCA	TATTACAAGC	2850
GCTATATCAG	CIGGIGCATG	TEGIESCITC	AGIATTITCT	GACCAGAGIA	2900
GAAGCGCAAC	TECACEIGIE	GGTTCCCCCC	CICAACGICC	GGGGGGGGG	2950
CGATGCCGTC	ATCTTACTCA	TGIGIGIAGI	ACACCCGACC	CIGGIATTIG	3000
ACATCACCAA	ACTACTCCIG	GCCATCITCG	GACCCCTTTG	GATTCTTCAA	3050
GCCAGTTTGC	TTAAAGICCC	CTACTICGIG	CGCGITICAAG	GCCLICICCC	3100
			AGGICATTAC		3150
			CCIAIGIGIA		3200
			CTGCGAGATC		3250
				ATCACGIGGG	
			TCAACGGCIT		3350
				GAATGGTCTC	
				CAGCAGACGA	
			TGACTGGCCG		3500
				AAACCTTCCT	
			TGICIACCAC		3600
			TCATCCAGAT		3650
				CCCCCTCATT	
			TIACCIGGIC		3750
CCGATGICAT	TCCCGIGCGC	CGGCGAGGTG	ATAGCAGGG	TAGCCIGCIT	3800

FIG. 4B

10 20 30 40 50	
1234567890 1234567890 1234567890 1234567890	
TOGOCCOGO COATTICCIA CITGAAAGOC TOCTOGGGG GICCGCIGIT	3850
GIGCCCCCC GCACACCCC TGGCCCIATT CAGGGCCGC GIGIGCACCC	3900
GIGGAGIGGC TAAAGCGGIG GACTITATCC CIGIGGAGAA CCTAGGGACA	3950
ACCATGAGAT CCCCGGIGIT CACGGACAAC TCCTCTCCAC CAGCAGIGCC	4000
CCAGAGCITC CAGGIGGOCC ACCIGCATGC TCCCACOGGC AGOGGIAAGA	4050
GCACCAAGGI CCCGGCIGCG TACGCAGCCC AGGGCIACAA GGIGIIGGIG	4100
CTCAACCCCT CIGITGCIGC AACGCIGGC TTTGGIGCTT ACATGICCAA	4150
GCCCATGGG GITGATCCTA ATATCAGGAC CGGGGTGAGA ACAATTACCA	4200
CIGGCAGOCC CATCACGIAC TOCACCIACG GCAAGITCCT TGCCGACGCC	4250
GGGIGCICAG GAGGIGCTTA TGACATAATA ATTTGIGAGG AGIGCCACIC	4300
CACGGATGCC ACATCCATCT TGGGCATCGG CACTGTCCTT GACCAAGCAG	4350
AGACTGOGGG GGCGAGACTG GTTGTGCTCG CCACTGCTAC CCCTCCGGGC	4400
TOCGICACIG TGICCCATCC TAACATCGAG GAGGITGCIC TGICCACCAC	4450
COCACACATC COCTITIACG GCAAGGCIAT CCCCCTCGAG GTGATCAAGG	4500
GOGGAGACA TOTCATOTTO TOCCACTOAA AGAAGAAGTG CGACGAGCTO	4550
GCCGCGAAGC TGGTCGCATT GCGCATCAAT GCCGTGGCCT ACTACCGCGG	4600
TCTTGACGIG TCTGTCATCC CGACCAGCGG CGATGTTGTC GTCGTGTCGA	4650
COGATGCICT CATGACIGGC TTTACOGGCG ACTICGACIC IGIGATAGAC	4700
TECAACACET GIGICACTCA GACAGTCGAT TICAGCCTTG ACCCTACCTT	4750
TACCATTGAG ACAACCACGC TCCCCCAGGA TGCTGTCTCC AGGACTCAAC	4800
GCCGGGGCAG GACTGGCAGG GGGAAGCCAG GCATCIATAG ATTTGTGGCA	4850
COGGGGAGC GCCCTCCGG CATGITCGAC TCGTCCGTCC TCTGTGAGTG	4900
CTATGACGCG GCCTGTGCTT GGTATGACCT CACGCCCGCC GAGACTACAG	4950
TIAGGCIACG AGCGIACATG AACACCCCGG GGCTTCCCGT GIGCCAGGAC	5000
CATCTIGAAT TITIGGGAGGG CGICTTTACG GGCCTCACTC ATATAGATGC	5050
CCACTITITIA TOCCAGACAA AGCAGAGIGG GGAGAACITT CCITACCIGG	5100
TACCITACCA ACCCACCGIG TGCCCTAGGG CTCAAGCCCC TCCCCCATCG	5150
TECENCAGA TETEGAAGIG TITGATCOC CITAAACCCA CCCICCATEG	5200
CCCAACACCC CICCIATACA CACIGGGGC TGITCAGAAT GAAGICACCC	5250
TGACGCACCC AATCACCAAA TACATCATGA CATGCATGTC GGCCGACCTG	5300
CACCICCICA CCACCACCIG GGICCICGIT GGCGGCGICC TGGCIGCICI	5350
GCCCCCTAT TGCCTGTCAA CAGCCTGCGT GGTCATAGTG GCCAGGATCG	5400
TOTTGICCGG GAAGCCGGCA ATTATTACCTG ACAGGGAGGT TCTCTTACCAG	5 4 50
CACTICCATG AGATGCAACA GIGCICICAG CACTIACCGT ACATCGAGCA	5500 5550
AGGGATGATG CICGCIGAGC AGIICAAGCA GAAGGCCCIC GGCCICCIGC	5550 5600
AGACCGCGTC CCGCCATGCA GAGGTTATCA CCCCTGCTGT CCAGACCAAC	5600 5650
TOGCAGAAAC TOGAGGICIT TIGGGCGAAG CACATGIGGA ATTICATCAG	5650 5700
TGGCATACAA TACTTGGCGG GCCTGTCAAC GCTGCCTGGT AACCCCGCCA	5/00

FIG. 4C

10 20 30 40 50 1234567890 1234567890 1234567890 1234567890 TTGCTTCATT GATGGCTTTT ACAGCTIGGG TCACCAGCC ACTAACCACT 5800 GCCCAAACC TCCTCTCAA CATATTGGG GGGGGGGG CTACCTGGG 5800 GCCCACAGCT GGGGGTGA CTGGGCTTTGT GGGGCGAGC CATAACCACT 5800 GCCCACAGC CAGGGTTGAA CTGGGCTTTGT GGGGCGAGC CTTACCTGGG 5800 GGGTAAGGGC CAGGGTTGAA CTGGGCTTTGT GGGCCAGCC CATTCTTGAC 5900 GGGTAAGGGC CAGGGTTGAA CTGGGCAGCT GAACCTGGGG CATTCTTGAC 5900 GGGTAAGGGC CAGCGTTGAA GGGGGGGG GGAACCTCTT GTACCATTCA ACATCACTGA 5950 GGGTAAGGGC CCCCCACGG AGGACCTCTT GTACCATTCA ACATCACTGA 5950 GGGTAAGGTC CCCTCACGG AGGACCTCTT GTACCATTCA ACATCACTGA 5950 GGGCACGTTG GCCCGACGG AGGACCTCTT GTACCATTCA ACATCACTGA 6000 TCTGGCCAGG ACCCTTGAG AGGACTGGT CAATCACTGC CAGCACTTCA ACATCACTGA 6000 ACCCTGGGAATGC ACCCCTCACG AGGACCTGGT CAATCACTGA CATACTGCGC 6050 ACCACGCTCC TCAGGGGGA ACCATGGTG AATCACCAG CCTCACTGTA 6200 ACCACGCTCC TCAGGGGGAC GCACTGCACA TACTCACCAG CCTCACTGTA 6200 ACCATGCTCC GGTTCCTGGC TAAGGGACAT CTGGGACTG AATTCCACCA 6250 TCCTGGCACTTC CTTTAACCAC TGGCTGCAAC CCTCACTGTA 6200 ACCATGCTCC GGTTCCTGGC TGAGGGACACT CTGGGACTG AATTCCACCA 6350 CCTGGGAATTC CCTTTTGTGC TGGCACGCC GGGTAATGGG GGGTCTGGGG 6400 AGGACACGGC ATTATCCACA CTGGCTGGCA CTGTGCACCA GGGCTCTGGGG 6550 CCCTGGGAATTC CCTTTGGGC ACTGTAACTT CCCCATGACCA GGGGCCTGGG 6550 CACATGTCCA AACGGGGGG ACTGTAAAGTT CCCCATGGCG ACCTTCACGG 6950 AACATGTGCG GGGGACACTT TAAGACCTGC GGGCACACCG ACCTTCACGG 6550 TACCCCCCTT CCTGGGCGAA ACTATAAGTT CCCCATGGC ACCTCCACGG 6600 CACAGGCAATC CTGCGGGGGG GAGGTATCAT TCACAGGTT GCGCACCGA 6700 ATTTTTCACCT CCTGCGGGGG GAGGTATCAT TCACAGGTT GCCCAGCC GGGGAAACGT 6750 CCACAGCCTT CCTGGGGGGG GAGGTATCAT TCACAGGTT GCCCAGCC GGGGAAACGT 6750 GCAAGCCTT CCTGGGGGGG GAGGTATCAT TCACAGGGTT GCCCAGCC GGGGAAACGT 6750 CCACAGCCTT CCTGGGGGGG TAACCACAC CTGCCCGGG GAGGTACCAT 7000 CTCCCCTGGCGGAGG TCACCCCAC TAGGACCGG CTCACCCGG GGGAAACGT 7000 CTCCCCTGGACAGGT TACAGGCC TCTCACAGGG GTCCACCAC GGGGAAACGT 7000 CTCCCCTGGACAGGT TACAGGCC TTTAGGGCGG GACCACACG GTCCACCAC 7350 AACCAGCCCTT CCTGGCGCAC TTCCACACA AAGTCTTGGC TTCCCCCCAC TCTCACCCAC ACCCC						
TRECTRICATE CATGGCTTTT ACACCTOCCG TCACCAGCCC ACTAACCACT 5750	10	20	30	-		
GSCCAAACCC TCCTCTCAA CATATTGGGG GGGTGGGTGG CTGCCTACCT CGCCGCCCCC GGTGCCCTTCAA CTGCCTTTGT GGGTGCTGGC CTACCTGGGG 5850 CCCCATCGG CAGCGTTGGA CTGGGGAGGT TCCTCGTGGA CATTCTTCCA 5900 GGGTGAGGTC CCCTCCAGGG AGGACCTCTT GTACCATTCA AGATCATCAG 5950 CGGTGAGGTC CCCTCCAGGG AGGACCTGGT CAATTCATCAG AGATCATCAG 6000 TCTGGCCTGG AGCCTTGTA GTGGGGTGG TCTGGGCACC AATTACTGGC 6000 TCTGGCCTGG AGCCCTTGTA GTGGGTGGT TCTGGCCACC AATTACTGGC 6050 CGGCAGGTTG CCCCGGGGAA ACCATGTTTC CCCCACCAC AATTACTGGC 6050 AGCCAGGTTC CCCGGGGGA ACCATGTTTC CCCCACCAC TAGGTGCGC 6150 AGCCAGGTCC TGAGGCCACT GCATCAGTGG ATTACCACCAC CTCCACTGTA 6200 ACCAGCTCC TGAGGCCACT GCATCAGTGG ATTACCACCAC CTCCACTGTA 6200 ACCAGCTCC TGAGGCCACT GCATCAGTGG ATTACCACC 6250 TCCATGCTCC GGTTCCTGGC TAGGGGACACTT CTGGCACCTC GCACACCTG 6300 TCCTGGGCAC CTTTTGTGGC CTGCCACCC GGGTCACACCTG GCACACCTG 6300 CCTGGGACTC CCTTTTGTGGC CTGCCACCC GGGTCACACCTG GCACACCTG 6300 CCTGGGACTC CCTTTTGTGGC CTGCCACCC GGGTCACACCTG GCACACCTG GCACACCCTT CCTGCCCCAC ACTTTACACTT CCCCCACTTCACACCA GCGGCCCCTT GCGCCCAA ACTATACACTA CCCCCACACCAC CAGGCCCCCTT GCGCCCAA ACTATACACCA GCGCCCCCTT GCGCCCAA ACTATACACCAC GCGCCCCCTT GCGCCCAA ACTATACACCAC GCGCCCCCTT GCGCCCAA ACTCCACACCTG GCGCCCCTT CCTCCACACCTG GCGCCCCTT CCTCCACACCTG GCGCCCCCTT GCGCCCAACCCCAA ACCCCCCCTT CCTCCACACCTC CACACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCC	<u>1234567890</u>					
CGCCGCCCC GGIGCGCIA CIGCTITIGI GGGIGCIGG CIACCIGGG 5850 CGGCATGGG CAGGGIGGC GGGACCICTI GTACCATTCA ACAICATCAG 5900 GGGIAGGGI CCCICCAGG AGACCICTI GTACCATTCA ACAICATCAG 5950 TCTICGCCIGG AGACCICTI GTACCATTCA ACAICATCAG 6900 TCTICGCCIGG ACACCIGGI GCGACGGIGIGG TCTICGCCACA ACATCCICTICG CCCCCACACCIC TACCIGGGG ACACCITCACC GCGCACGGIG CAATCCICTICA ACCCCTACAC TACCIGGGGG 6150 AGACCATIGC TCCCGGGGGA ACCATGITTC CCCCACGCAC TACCIGGGGG 6150 AGACCATICC TCAGGGGGAC GCATCACTGCA ATACCTCAGG CCTCACTGGA 6200 ACCCAGCICC TGAGGGGACT GCATCAGGG ATACCTCAG ACTGGGGACAC TCCATGGGG ACACCACTG GCACCACTG GACCACCTC GGGGCACCT GAGGGGACACT GCATCAGGG ATACCTCG ACTGGGACAC TCCATGGCAC TCCAGGCAC TGAGGGGACACT GCATCAGGG ATACCTCG ACTGGGACAC TCCAGGCAC TCCATGGGACAC TCCAGGCAC TGAGGGGACACTG GAGACACCG CCTCACTGGA ACACCACTCC GGGGACACTG GAGACACCG CAGACCTCC GAGACACCG GGGGACACTG GAGACACCAC TCCATGGACAC CTCACTGCAC CAGACCTCAT GCCACAACTG 6350 CCTGGGACTC CCTTTGGGCC ACCCCACTGACAC CGGGCACACTG GACACCTG CAGACCTG GACACCTG GACACCCTT CCCCACTGCAC ACCTGCACAC GACACCCTG GACACCCTT CCTGCCCCA ACTATAACGCT TCCACACACCT GACACCCCTG GACACCCCTT CCTGCCCCA ACTATAACGCT TCCACACACCT GACACCCCTT CCTGCCCCA ACTATAACGCC TCCACACCTG GACCTCACCAC GACACCCCTT CCTGCCCCA ACTATAACCAC CACACCGTT GACACCCCTT TCCACACCAC GACCCCCTT GACACCCCTT TCCACACCAC GACCCCCCT GACCACCTT TCCACACCAC CACCCCCTT GACCCCCTT TCCACACCAC GACCCCCCCACCCG GACCACCCCC ACCCCACCCC ACCCCACCCCC ACCCCACCCC ACCCCCC	TIGCTICATT	GATGGCTTTT	ACAGCTGCCG	TCACCAGCCC	ACTAACCACT	5750
CCCCATCGG CAGGITIGA CIGGGAAGG TOCTOGIGGA CATTOTICA 5900 GGGIATIGAG GGGAACTOTT GIACCATTCA AGAITATICA 5950 CGGIATIGAG GGGAACTOTT GIACCATTCA AGAITATICAG 5950 CGGIAGGTI CCCTCAAGG AGAACTIGGT CAATCTGTIG CAGCATTCA AGAITATICAG 6000 TCTCGCCTIGG AGCCTTIGIA GIOGGIGTG TCTGCGCAC AATTACTCAGC 6050 CGGCAAGTTIG CCCCGAGGAA GGGGGCAGTIG CAATCTGTATA ACCGCTATAT 6100 AGCCTTCCC TCCCGGGGAA ACCATGITTC CCCCAGGAC TAGGICAGG 6150 ACACCGCTCC TGAGGGCACT GCATCAGTGA ATAACTCGG AGTGTACAC 6250 TCCATGCTCC GGTTCCTGGC TAAGGGCACA TAGTCAGCAG ATAGTCAGGG 6250 TCCTGGGATTC CTTTAGGAC TGACTCAGTGA ATAACTCGG ATATGCACG 6250 TCCTGGGATTC CCTTTGTGTC CTAAGGGACAT CTGGGACTGA ATATGCAGG 6250 CCTGGGATTC CCTTTGTGTC CTGCCTGAAGA CTGGGACTGA ATATGCAGG 6350 CCTGGGATTC CCTTTGTGTC CTGCCAGGC GGGTATAGG GGGTCTGGG 6400 AGCACGCTC CTTTTAGGAC TGGCTGCACA CTGGGACTGA GACATCACTG 6900 GACATGTCAA AAACGGAACG ACCATCACACA CTGGGACTCATG GACATCACTG 6900 GACATGTCAA AAACGGAACG ACCATTAAACGT TCGGTCCTAG GACATCACTG 6950 AACATGTGGA GTGGGACATA AGCGGGTG GGGACTCCACACA CGGGCCCCTG 6550 TACTCCCCTT CCTGCCCCA ACTATAACGTT CCCCTTGTGG CACCTCACGG 6650 CAGAGGAATA CGTGGACAATC TAAATGCCC TGCCCACATCC CACCGCTG 6550 ACTTTTACACA CAATTGCACAG GGGGGGCCT ACCACATCC CACCGCTG 6650 GGTATGACTAC CTGCCCCAAACTC TAAATGCCG TGCCCACATCC CACCGCTG 6650 GGTTGACGTC CTGCCCCAAACTC TAAATGCCC TACCACATCC CACCGCCG 6650 TCCCCGGGGG GGGCCCTA ACCACAGGTTT CCCCCTTTCACGG 6660 TCCCCGGGGG GGGGACTCA TACCACACGG ACCTCACCGG 6660 TCCCCCTTCACGC CACCACCTT TAAATGCCC TCCCCCCTT CACCACCCG CCCCCCCCCC	GGCCAAACCC	TCCTCTTCAA	CATATIGGG	GGGIGGGIGG	CIGCCCAGCT	5800
GGGTATIGGG GGGGGTGGC GGGAGCTCTT GTAGCATTCA AGATCATGAG 5950 GGGTCAGGTC COCTICAGGG AGGACCTGGT CAATCTGCTG GGGGCATCC 6000 TCTCGGCTGG AGGCCTTTGTA GTGGGTGTG CAATCTGCTG GGGGCATCC 6050 GGGCAGGTTG GCCGGGGGAGA AGGACCTGGT CAATCTGCAC AATACTGCGC 6050 AGGACCTTGC CGCGGGGAAAAAAAAAAAAAAAAAAAAA	æææææ	GGIGCCGCIA	CIGCCITIGI	GGGIGCIGGC	CIACCIGGG	5850
CGGTCAGGTC CCCTCCAGGS AGGACCTGGT CAATCTGCTG CCGGCATCC 6000 TCTCGCCTGS AGCCCTIGTA GTCGGTGTGG TCTGGCCAGC AATACTGGCC 6050 CGGCAGGTTG GCCGGGGAA ACCATGTTTC CCCCAGGCAC TACGTGCGGG 6150 AGCCTTCGCC TCCGGGGGAA ACCATGTTTC CCCCAGGCAC TACGTGCGGG 6150 AGCCAGCTC TGAGGGGCAA ACCATGTTTC CCCCAGGCAC TACGTGCGGG 6200 ACCCAGCTCC TGAGGGGCAA CACATGTTCC CCCACGCAC CTCACTGTGG 6200 ACCCAGCTCC TGAGGGCACT GCATCAGTGG ATACCCCAGG ACTCACTGTG 6200 TCCATGCTCC GGTTCCTGGC TAAGGGCACT CTGGGACTGG ATATCCCAGG 6300 TCCTGGGATTC CCTTTGTGTC CTGCCAGGC GGGTATAGGG GGGTCTGGG 6350 CCTGGGATTC CCTTTGTGTC CTGCCAGGC GGGTATAGGG GGGTCTGGG 6400 AGGACACGC ATTATGGACA CTCGCTGCAC CTGTGGAACT GACAACCTG 6950 CACATGTCGA AAACGGCACA CTGCCTGCACC CTGTGGAACT GACATCACTG 6950 AACATGTGCA GTGGGACGTT CCCCATTAAC CCCTTACACCA GGGCCCCTTG GGGACCCTTG 6650 TACTCCCCTT CCTGCGGCCA ACTATAAGGTT CGCGCTGTGG AGGGTGTTCTG 6600 CACAGGCAATA CGTGCACATA AGGCGGGTGG GGGACTTCCA CTACGTATCG 6650 CGTATGACTA CTGCACAATCT TAAATGCCG TGCCACACTC CATCGCCCAC 6650 GGTATGACTA CTGCACAATCT TAAATGCCG TGCCACACTC CATCGCCCAC 6700 ATTTTTTCACA GAATTTGAGG GGGTGCCCT ACAACAGGTTT GCGCCCCTT 6750 CCACCCCTTT CCTGCGGCAC GAGGTTACCCT ACAACAGGTTT GCGCCCCTT 6750 CCACCCCTTT CCTGCGGAC GAGGTTACCCT ACAACAGGTTT GCGCCCCTT 6750 CGACACCCTTT CCTCCGGACG GAGGTTACACCA ACTCCACGGG 6800 CGACAACGTTT GCTCCCACATT ACCCTCCCA TATTAACACCA GAGGCGGCC 6900 CGACAACGTTT GCCCCCATC TCCACAGGAC CTCCTCGGCT 6950 CGACAACGTTT GCCCCCATC TCCACAGGAC CTCCTCGGCT 6950 CGCGCACACCTTT CCCCCCTCCA TATTAACACCA GAGGCGGCC 6900 CGACAACGTTC ATCCCCCCTT CTATGGCAG CCCAACCATCA 7000 CCCCCCTGAC GCCCCCTTC TCCACGCAG ACGTACACATCA 7000 CCCCCCTGAC ACCCCCTT CTATGGCAG CTCCTCCGGT 7100 CCCCCCTGAC ACCCCCTT CTATGCCAG CACCACTGC 7100 CCCCCCCTCCACAA ACTTCCCCCC TTCCGCCAC GGTCAACCAG GGTCAACAAAAG CGTCCTCCCG T7100 CCCCCCCCCCT GTCGCCCCCC TTCGCCACAC ACCATCCCCC 7300 CCTCCCCTGAC ACCCCCCTT CTCGCCACA AAGTTCTCCGC 7300 GTCCCCTCCT GTCGCCCCCC CTCCCCACCT CCCCCCACTC CACCTCCAC 7300 CCCCCCCCC TTCGCCCAC TTCGCCCAC CCCCCCCCCC	CCCCCATCCC	CAGCGITGGA	CIGGGGAAGG	TCCTCGTGGA	CATTCTTGCA	5900
TCTCCCCTICS ACCCTTIGIA GTCGGTGTGG TCTCCCCACC AATACTCCCC GOSCACGTTG GCCGGCACGTTG GCCGGCACGTTG GCCGGCACGTTG CAATGCATCA ACCCGCTCACTGTA GCCGACCTC TCACGGCACC TCCCTGTA GCCACCCCC TCACGGCAC ACCCCACCCC	GEGITATIGECG	CEGCCIGCC	GGGAGCICIT	GIAGCATICA	ACATCATGAG	5950
CGGCACGTTG GCCCGGGGA GGGGGCAGTG CAATGGATCA ACCGCTAGTA 6100 AGCCTTCGCC TCCCGGGGA ACCATGTTTC CCCCACGCAC TACGTGCGG 6150 AGAGCGATGC AGCGCCCC GTCACTGCCA TACTCAGCAG CCTCACTGTA 6200 ACCCACCTCC TGAGGCCACT GCATCAGTGG ATACCTCGGA AGTGTACCAC 6250 TCCATGCTCC GGTTCCTGGC TAAGGGCACAT CTGGGACTGG ATATGCCACG 6350 TGCTGAGCGA CTTTAAGCACC TGGCTGAAAG CCAAGCTCAT GCCACACTGG ATATGCCACG 6350 CCTGGGATTC CCTTTGTGTC CTGCCAGCGC GGGTATTAGGG GGGTCTGGGG 6400 AGGACACGGC ATTATGCACAC CTGCCTGCAA CTGTGGACCT GACATCACTG 6900 GACATGTCAA AAACGGGACG ATGAGGATCG TCGGTCCTAG GACATCACTG 6900 GACATGTGAA AAACGGGACG ATGAGGATCG TCGGTCCTAG GACATCACTG 6950 AACATGTGGA GTGGGACGTT CCCCATTAAC GCCTTACACCA CGGGCCCCTG 6550 TACTCCCCTT CCTGGGCCGA ACTATAAGGTT CGCCTTGTGG AGGGTGTCTG 6660 CAGAGCAATA CGTGCACATA ACCGGGTGG GGACTTCA CTACGTATCG 6650 GTTATCACTA CTGACAATCA TAAATGCCCG TCCCAGTTCA CTACGTATCG 6650 GTTATCACTA CGACACATA ACCGGGTGG GGACTTTCA CTACGTATCG 6650 GTTACCCGTTG CTGCGCGCAG GAGGTGCGCT ACACAGGTTT GCCCCCCCTT 6750 GCAAGCCCTT CCTGCGCGAG GAGGTACCAT TCACGTAGG ACCGTACCCG 6660 TACCCGGTGG GGTCGCCAT ACCCAGTACC ACCCAGTCC CATCGCCCCA 6700 ATTTTTCACA GAATTGCACG GAGGTGCGCT ACACAGGTTT GCCCCCCCTT 6750 GCAAGCCCTT CTGCGCGAGG GAGGTACCAT TCACAGTAGG ACCTCACCAG 6800 TACCCGGTGG GGTCGCCAATT ACCTTGCGAG CCCGAACCG ACGTAGCCGT 6850 GTTGACGTCC ATGCTCACTT TCTCAAGCAC ACTTGCACCAG GAGGCGCCG 6900 GAAGAAGGTT GCCCCCCAATCT TCTCAAGCAC ACCTTGCACC CCAACCATCA 7000 CCACCCCTCAC CCCCCAACCAG GTTCAGCCCT TCTCAAGCAC ACCTTCACCAC CCAACCATCA 7000 CCCCCCCCCCACCTC TCTCAAGACA ACTTCCCCTGTGG AGCACACATCA 7000 CCCCCCCCCCC CCCCAACCAG GTTCAGTCCC TTCCCCTGCC 7300 TCTGGGCGC CCCGAACCAG GTTCAGTCCC TTCCCCTCACC 7300 TCTGGCCCCCC GCCCCAACCAG GTTCAGTCCC TTCCCCCCCCCC	CCCTCACCTC	CCCTCCACCG	AGGACCIGGT	CAATCIGCIG	CCCCCATCC	6000
AGCOTTOGOC TOCOGGGGGA ACCATIGITTIC COCCAGGCAC TAGGIGGGGG 6150 AGAGCGATIC AGCOGCCGC GICACTIGCA TACTTCAGCAG COTCACTIGTA 6200 ACCCAGCTOC TIGAGGGGACT GCATCAGTIGS ATAAGCTGGG AGTIGTACCAC 6250 TOCATGCTCC GGTTCCTGGC TAAGGGCACT CTGGGGACTGG ATATGGGGAGG 6300 TGCTGAGGGC CTTTGAGACC TGCCTGAAG CCAACCTCAT GCCACAACTGG 6350 CCTGGGATTC CCTTTGTGTC CTGCCAGGGC GGGTATAGG GGGTCTGGGG 6400 AGGAGAGGGAC ATTATGGCACA CTGGTGCCA CTGTGGGAGCT GACATCACTG 6900 GACATGTCAA AAACGGGACG ATGAGGATTG TCGGTCCTAG GACCTGCAGG 6950 AACATGTGCA ATTATGCACA CTCGCTGCCA CTGTGGGACT GACATCACTG 6950 AACATGTGCA GTGGGACGTT CCCCATTAAC GCCTACACCA CGGGCCCTTG 6550 TACTCCCCTT CCTGGGCGA ACTATAAGGTT GCGCTGTGG AGGGTGTCTG 6600 CAGAGCAATTA CGTGGACAATA AGGCGGTTG GGCACTTCCA CTACGTATCG 6650 GTTATGACTA CTGACAATCT TAAATGCCCT GCCAGATCC CATCGCCCGA 6700 ATTTTTCACA GAATTGCAGG GGGGGGCCT ACACAGGTTT GCGCCCCTT 6750 GCAAGCCCTT GCTGGGGGGG GAGGTGCAT TCACAGGTTT GCGCCCCTT 6750 GCAAGCCCTT GCTGGGGGGG GAGGTGCAT TCACAGGTGG ACGTGACCGT 6850 TTACCCGGTGG GGTCGCAATT ACCTTGCGGG CCCGAACCG ACGTAGCCGT 6850 GTTGACGGTC ATGCTCACTG ATCCTTCCCA TATAACACCA GAGGGGGCG 6900 GGAGAAGGTT GCCGCAACTG ATCCTCCCA TATAACACCA GAGGGGCCG 6900 GGAGAAGGTT GCCGCACACTG TCCCAACCG ACCTGCGCT 6950 AGCCACCTGT CCCTCCCAT TACAAGGCA ACTTGCACGG CCAACCATCA 7000 CTCCCCTGAC GCCGAGCTCA TACAGGCGA ACTTGCACGG CCCAACCATCA 7000 CTCCCCTGCA CCCGAGCTCA TACAGGCGA ACTTGCACGG GCCAACCATCA 7000 CTCCCCTGCA ACCCCCTT CTCCAAGGCA ACTTGCCCCG 7100 CACTCCCTTCG ATCCGCTTCTC GGTCCATCGC TCCCTGGGG AGCCCTCCCG 7200 TCTGGGGGGG CCGGGACTCA TACAGGCGA ACTTGCCCGG GCCCACCATCAC 7100 GACTCCCTTCG ATCCGCTTCTC GGTCCATCGC TCCCTGGGG AGCCTCCCGC 7200 TCTGGGCGCG CCCGAACCAC ACCTCCAC GGTCAACCATCA 7250 CCTGCACCCT GTCCCCCCCCCCCCCCCCCCCCCCCCCCC	TCTCCCCTCG	AGCCCTTGTA	GICGGIGIGG	TCTGCGCAGC	AATACTGCGC	6050
ACAGCATICA AGCOCCCCC GITACTICCA TACTCACCAG CCTCACTIGITA 6200 ACCCAGCTIC TGAGGCCACT GCATCAGTIGS ATTACCTCCG AGTIGITACCAC 6250 TCCATCCTICC GGITTCTIGGC TAAGGCACAT CTGGGACTIGG ATTATCCCAGG 6300 TGCTGAGCCA CTTTAAGACC TGGCTCAAAG CCAACCTCAT GCCACAACTIG 6350 CCTGGCATTC CCTTTGTGTC CTGCCACCC GGGTATAGGG GGGTCTGCGG 6400 AGGAGACGCC ATTATCACAC CTGCCTCCCA CTGTGCACCT GACATCCTG 6900 GACATGTICAA AAACGCCACA ATTATCACAC ATTCACCACA CTGCCACCT GACATCCAGG 6950 AACATGTICGA ATTATCACA ATTCATCACAC ATCCCTCTAG GACCTCTG 6550 TACTCCCCTTT CCTGCCCCCAA ACTTATAAGTT CCCCCTTGTGG AGGGTGTCTG 6600 CACAGGAATA CGTGCACATA AGCCCGGTGG GACCTTTCA CTACGTATCG 6650 CGTATGACTA CTGCACAATCT TAAATCCCC TGCCACATC CACCGCCCA 6700 ATTTTTCACA GAATTGGACG GGGTGCCCT ACACAGGTTT GCCCCCCTA 6750 CCAAGCCCTT GCTGCCGAAGT ACCTTCCACCAGG ACTTACACCA GACTTACCACGAG 6800 TACCCGGTGG GGTCCCAATT ACCTTCCCA TACAACACCG ACGTACCCGT 6850 GTTCACGGTC ATCCTCCACT ATCCTCCCA TATAACACCA GACGCCCG 6900 GCAGAACGTT GCCCCAACCT TCTCAACGAA ACTTCCACCAG G6900 GCAGAACGTT CCCCCCACCTC TCTCAACGAA ACTTCCACCAG GACGCCCG 6950 CCCCCCCCCCACCCG CCCCAACCG CTCCTCCGCC TCCCCCCCCC TCTCCCCCT TCTCCACCACG CTCCTCCCCT TCTCCACCACG CTCCTCCCCT TCTCCACCACG CTCCTCCCCT TCTCCACCACG CTCCTCCCCT TCTCCACCACCACC CTCCTCCCCT TCTCCACCACCACCACCACCACCACCACCACCACCACCAC	CCCCACCTIC	GCCCGGGGGA	GGGGGCAGIG	CAATGGATGA	ACCGCTAAT	6100
ACCCACCTIC TGAGGGGACT GCATCAGTGG ATAACCTGG AGTGTACCAC 6250 TCCATGCTCC GGTTCCTGGC TAAGGGACAT CTGGGACTGG ATATGCGAGG 6300 TGCTGAGCGA CTTTAAGACC TGGCTGAAAG CCAAGCTCAT GCCACAACTG 6350 CCTGGGATTC CCTTTGTGTC CTGCCAGGG GGGTATAGG GGGTCTGGG 6400 AGGAGACGGC ATTATGCACA CTGGCTGCCA CTGTGGAACT GAGATCACTG 6900 GACATGTCAA AAACGGGACGT CCCCATTAAC GCCTACACCA CGGGCCCCTG 6550 AACATGTGGA GTGGGACGTT CCCCATTAAC GCCTACACCA CGGGCCCCTG 6550 TACTCCCCTT CCTGCGCGA ACTATAAGTT CGCGCTGTGG AGGGTGTTG 6600 CACAGGAATA GGTGGACATA AGGGGGGTGG GGGACTTTCCA CTACGTATCG 6650 GGTATGACTA CTGACAATCT TAAATGCCG TGCCACATCC CATCGCCGA 6700 ATTTTTTCACA GAATTGGACG GGGTGGGCCT ACACAGGTTT GCGCCCCTT 6750 GCAAGCCCTT GCTGCGGAGG GAGGTGCCCA TACACAGGTTT GCGCCCCCT 6750 TACCCGGTGG GGTCGCAATT ACCTTGCGAG CCCGAACCG ACGTACCCG 6690 TACCCGGTGG GGTCGCAATT ACCTTGCGAG CCCGAACCG ACGTACCGT 6850 GTTGACGTCC ATGCTCACTG ATCCTCCA TATAACAGCA CAGGGGCCG 6900 GGACAAGGTT GCCGCACATC TCCCACAGG CTCCTCGGCT 6950 AGCCAGGTGT CCGCTCCATC TCCCAAGCG ACGTACCGT 6950 AGCCAGCTTG CCGCTCCATC TCCCAAGCG ACGTACCGT 6950 CTCCCCTGAC GCCGACCTCA TACACGCTA CCTCCTGTGG AGGCACGAGA 7000 CTCCCCTGAC GCCGACCTCA TACACGCTA CCTCCTGTGG AGGCACGAGA 7050 TCGGGGGAA ATTCTCCGGA AGTCTCCGA ACTCCCCGT CTATGCCCG CCAACCATGA 7000 CTCCCCTGAC GCCGACCTCA TACACGCTA CCTCCTGTGG AGGCACACGA 7050 TCGGGGGACTCA TACACGCTA CACCCCCCT CTATGCACG GCCTGACACG 7200 TCTGGGGGGG GCCGACTCA TACACGCTA ACCCCCCC TAGTACACG GTGGAAAAAG 7250 CCTGACTACG AACCACCTGT GTCCATGGA ATTCCCCCG CACCCACCGG 7350 GTCCCTCCT GTGCCTCCGC CTCGGAAAAA GGTTACCGGG ACGTCCACG 7350 AATCAACCTT ATCTACTGCC CTCGGAAAAA GGTTACCGGG 7350 AATCAACCTT ATCTACTGCC TTGGCCGAC TTCCCCCCA 7450 AACCACCCCCC TTCTGCCCAC TTCGCCCAC TACTTCCACA AAGTTTTCGC AACCACCTCT TCTGCCCCAC TTCGCCACAA AAGTTTTCGCCAA AATCCACCAC TTCGCCCCCC TTCGCACAAA AAGTTTTCGCCAA AATCCACCAC TTCGCCCCCC TTCGCCCCCC TTCGCCCCCC TTCGCCCCCC TTCGCCCCCCCC	AGCCTTCGCC	TCCCGGGGGA	ACCATGITIC	CCCCACGCAC	TACGIGCCGG	6150
TOCATOCTIC GGITTOCTIGGC TAAGGGACAT CTGGGACTIGG ATATGGGAGG 6300 TIGCTGAGGGA CTTTAAGACC TGGCTGAAAG CCAAGCTCAT GCCACAACTG 6350 CCTGGGATTC CCTTTGTIGTC CTGCCAGGGC GGGTATAGGG GGGTCTGGGG 6400 AGGAGACGGC ATTATGCACA CTGGCTGCCA CTGTGCAGCT GACATCACTG 6900 GACATGTCAA AAACGGGACG ATGAGGATCG TGGGTCCTAG GACATCACTG 6900 GACATGTCAA AAACGGGACG ATGAGGATCG TGGGTCCTAG GACATCACTG 6950 AACATGTGCA GTGGGACGTT CCCCATTAAC GCCTACACCA CGGGCCCCTG 6550 TACTCCCCTT CCTGGGCCCA ACTATAACTT CGCCCTGTGG AGGTGTCTG 6600 CAGAGGAATA CGTGGAGATA AGGGGGGGGGGGGGGGGG	AGAGCGATGC	AGCCGCCCCCC	GICACIGOCA	TACTCAGCAG	CCTCACTGIA	6200
TGCTGAGCGA CITTRAGACC TGGCTGAAAG CCAAGCTCAT GCCACAACTG 6350 CCTGGGATTC CCTTTGTGC CTGCCAGCG GGGTATAGGG GGGTCTGGGG 6400 AGGACACGGC ATTATGCACA CTGCTGCCA CTGTGCAGCT GACATCACTG 6900 GACATGTCAA AAACGGACG ATGAGGATCG TCGGTCCTAG GACCTGCAGG 6950 AACATGTGGA GTGGGACGTT CCCCATTAAC GCCTACACCA CGGCCCCTG 6550 TACTCCCCTT CCTGCGCCCA ACTATAAGTT CGCCCTGTGG AGGGTGTGG 6600 CACAGGAATA CGTGGACATCA AGGCGGTGG GGGACTTCCA CTACGTATCG 6600 CACAGGAATA CGTGGACATCA TAAATGCCCG TGCCAGATCC CATCGCCCA 6700 ATTTTTCACA GAATTGGACG GGGTGCGCCT ACACAGGTTT GCGCCCCCTT 6750 GCAAGCCCTT GCTGCGGGAG GAGGTATCAT TCAGAGTAGG ACTCCACGAG 6800 TACCCGGTGG GGTCGCAATT ACCTTGCCAG CCCGAACCG ACGTAGCCGT 6850 GTTGACGTCC ATGCTCACTG ATCCCTCCCA TATAACAGCA GAGCGGCCG 6900 GCACAAGGTT GCCGCCAACTT TCTCAAGGCA ACTTGCCAGG CTCCTCGGCT 6950 AGCCAGCTGT CCGCTCCATC TCTCAAGGCA ACTTGCACGA CTCCTCGGCT 6950 ACCCAGCTGT CCGCTCCATC TCTCAAGGCA ACTTGCACGG CCAACCATCA 7000 CTCCCCTCGC CCCGACCTCA TAGCAGCGTA ACACAAAGT GGTGATTCTG 7100 GACTCCTTCG ATCCCTCTCT TGCAGGCA ACTTGCCCCG CCAACCATCA 7000 CTCCCCTCGC CCGCACCTC TGCAGAGCGA ACTTGCCCCG CCCAACCATCA 7000 CTCCCCTCGC GCCGACCTC TGCAGAGCGA ACTTGCCCCG CCCAACCATCA 7000 CTCCCCTCGC GCCGACCTC TGCAGAGCGA ACTTCCCCCG T750 ACCTGCACGAA ATTCTGCGCA AGTCTCCGCA ACTCACCCG CTCCCCCCG T750 CCTCCCCTCCA GCCGACTCC TCCGCACACCA CATCACCCCC TTCGCCCCCC TTCGCCCCCCC TTCGCCCCCCC TTCGCCCCCCC TTCGCCCCCCC TTCGCCCCCCC TTCGCCACCA AACCAACAA ACGTTTTCGCC CCCCCCACC TTCCCCCCCCCC	ACCCAGCTCC	TGAGGCGACT	GCATCAGIGG	ATAAGCTCGG	AGIGIACCAC	6250
CCTGGGATTC CCTTTGTGTC CTGCCAGGGC GGGTATAGGG GGGTCTGGGG 6400 AGGAGAGGGC ATTATGCACA CTGGCTGCCA CTGTGGAGCT GAGATCACTG 6900 GACATGTCAA AAACGGGAGG ATGAGGATCG TGGGTCCTAG GACCTGCAGG 6950 AACATGTGGA GTGGGACGTT CCCCATTAAC GCCTACACCA CGGGCCCCTG 6550 TACTCCCCTT CCTGCGCCGA ACTATAAGTT CGCGCTGTGG AGGGTGTCTG 6600 CAGAGGAATA CGTGGAGATA AGGCGGTGG GGGACTTCCA CTACGTATCG 6650 GGTATGACTA CTGACAATCT TAAATGCCCG TGCCAGATCC CATCGCCCGA 6700 ATTTTTCACA GAATTGGACG GGGTGGGCCT ACACAGGTTT GCGCCCCTT 6750 GCAAGCCCTT GCTGCGGGAG GAGGTATCAT TCAGAGTAGG ACTCCACGAG 6800 TACCCGGTGG GGTGCCCAATT ACCTTGCCGA CCCGAACCG ACGTAGCCGT 6850 GTTGACGTCC ATGCTCACTG ATCCCTCCCA TATAACACCA GAGGGGCCG 6900 GGACAAGGTT GCGCACAGG TCACCCCCTT CTATGGCCAG CTCCTCGGCT 6950 AGCCACCTGT CCGCTCCATC TCTCAAGGCA ACTTGCACG CCAACCATGA 7000 CTCCCCTGAC GCCGACCTGA TAGAGGCTAA ACTTGCACGG ACGTCACTTC 7100 GACTTCTCTGA GCCGCACCTG TTCAAGGCA ACTTGCACGG ACGTCACTTC 7100 GACTTCTCCTGAC GCCGACCTGA TAGAGGCTAA ACTTGCACGG ACGTCACTTC 7100 GACTGCCTGAC ACTCACCAGG GTTCAGGCTAACAAAAGT GGTGATTCTG 7100 CACTGCACAA ATTGTGCGCA AGTCTCGAGG ACGACAACAT 7250 CCTGCACTAA ATTGTGCGCA AGTCTCGAG ACACAACAT GGTGCACTACA 7250 CCTGCACTAA ATTGTGCGCA AGTCTCGAGC TTCCCCCTTAC CACCTCCACG 7300 GTCCCCTCCT GTGCCTCCCC CTCGGAAAAAA GCGTTACCGTG GTCCATCACCG 7350 AATCAACCCT ATCTACTCCC TTGGCCGACC TTCCCCCTAC CACCTCCACG 7350 AATCAACCCT ATCTACTCCC TTGGCCGAC TTGCCACCAA AAGTTTTGGC 7400 ACCTCCTCAA CTTCCCCCAT TACGGGCGAC TTCCCCCTAC AAGTTTTTGC 7400 ACCTCCTCAA CTTCCCCCCT TTGGCCGAC TTCCCCCCTAC AAGTTTTTGC 7400 ACCTCCTCAA CTTCCCCCAT TACGGGCGAC TTCCCCCCAAAAAAAAAA	TCCATGCTCC	GGITCCIGGC	TAAGGGACAT	CIGGGACIGG	ATATGCGAGG	6300
AGGAGACCIC ATTATOCACA CTCCCTGCCA CTGTGGAGCT GAGATCACTG 6900 GACATGTCAA AAACGGACG ATGAGGATCG TCGGTCCTAG GACCTGCAGG 6950 AACATGTGCA GTGGGACGTT CCCCATTAAC GCCTACACCA CGGGCCCCTG 6550 TACTCCCCTT CCTGCGCCGA ACTATAAGTT CGCGCTGTGG AGGGTGTCTG 6600 CAGAGGAATA CGTGGAGATA AGGCGGGTGG GGGACTTCCA CTACGTATCG 6650 GGTATGACTA CTGACAATCT TAAATGCCCG TGCCAGATCC CATGGCCCGA 6700 ATTTTTCACA GAATTGGACG GGGTGCGCCT ACACAGGTTT GCGCCCCTT 6750 GCAAGCCCTT GCTGCGGGAG GAGGTATCAT TCACAGTAGG ACTCCACGAG 6800 TACCCGGTGG GGTCGCAATT ACCTTGCCAG CCCCAACCGG ACGTAGCCGT 6850 GTTGACGTCC ATGCTCACTG ATCCCTCCCA TATAACAGCA GAGGGGGCG 6900 GGAGAAGGTT GCCGACAGGG TCACCCCCTT CTATGGCCAG CTCCTCGGCT 6950 AGCCAGCTGT CCGCTCCATC TCTCAAGGCA ACTTGCACGG CCCAACCATGA 7000 CTCCCCTGAC GCCGAGCTCA TAGAGGCTAA CCTCCTGTGG AGGCAGGACA 7050 TGGGCGGCAA CATCACCAG GTTGAGTCAG ACAACAAAGT GGTGATTCTG 7100 GACTCCTTCG ATCCCTTGT GCCAGAGGG GATGACCAGG GATGACCAGG 7200 TCTGGCGCAAA ATTCTGCGGA AGTCTCGGAG GATGACCAGG GCCTGCCCG 7200 TCTGGCCGAAA ATTCTGCGGA AGTCTCGGAG GATGACCAGG GCCTGCCCG 7300 GTCCCCTCCT GTGCCTCCGC CTCGGAAAAA GCGTACGGTG GTCCTCACCG 7350 AATCAACCCT GTGCCTCCGC TTGGCCGACC TTGCCACCAA AAGTTTTGGC 7400 AGCTCCTCAA CTTCCGGCATT TACGGGCGAC TTGCCACCAA AAGTTTTTGGC 7400 AGCTCCTCAA CTTCCGGCATT TACGGGCGAC AATCACCAAA AAGTTTTTGGC 7400 AGCTCCTCAA CTTCCGGCATT TACGGGCGAC AATCACCAAA AAGTTTTTGGC 7400 AGCTCCTCAA CTTCCGGCATT TACGGGGCACCAC AATCACCAAA AAGTTTTTGGC 7400 AGCTCCTCAA CTTCCGGCATT TACGGGGCAC TTGCCACCAA AAGTTTTTGGC 7400 AGCTCCTCAA CTTCCGGCATT TACGGGGCAC TCCCACGT TCCCACCTAC TCCTATTCTT 7500 CCATGCCCCCCT TCTGGCCACCT CCCCCCGACTC CACCGCACTTC CACCCCCCC 7450	TGCTGAGCGA	CITTAAGACC	TGGCTGAAAG	CCAAGCTCAT	GCCACAACTG	6350
GACATGICAA AAACGGACG ATGAGGATCG TOGGICCTAG GACCTGCAGG 6950 AACATGIGGA GIGGGACGIT COCCATTAAC GCCTACACCA CGGGCCCCTG 6550 TACTCCCCTT CCTGCGCCGA ACTATAAGIT CGCGCTGIGG AGGGIGTCTG 6600 CACAGGAATIA CGTGGAGATA AGGCGGGTGG GGGACTTCCA CTACGITATCG 6650 GGTATGACTA CTCACAATCT TAAATGCCCG TGCCAGATCC CATGGCCCCA 6700 ATTITITCACA GAATTGGACG GGGTGCGCCT ACACAGGITT GCGCCCCCTT 6750 GCAAGCCCTT GCTGCGGGAG GAGGTATCAT TCACAGTAGG ACTCCACGAG 6800 TACCCGGTGG GGTCGCAATT ACCTTGCCGAG CCCCAAACGG ACGTAGCCGT 6850 GTTGACGTCC ATGCTCACTG ATCCCTCCCA TATAACAGCA GAGGCGCCT 6950 GGAGAAGGTT GCCGAGAGGG TCACCCCCTT CTATGGCCAG CTCCTCGGCT 6950 AGCCAGCTGT CCGCTCCATC TCTCAAGGCA ACTTGCACG CCAACCATGA 7000 CTCCCCTGAC GCCGAGCTCA TAGAGGCTAA CCTCCTGTGG AGGCAGGAGA 7050 TGGGCGGCAA CATCACCAG GTTGAGTCAG ACAACAAAGT GGTGATTCTG 7100 GACTCCTTCG ATCCGCTTGT GCCAGAGGAG GATGACCAGG ACGTACCATGA 7050 TCTGGCGCAAA ATTCTGCGGA AGTCTCGGAG GATGACCAG GCCCTGCCCG 7200 TCTGGGCGCAA ATTCTGCGGA AGTCTCGGAG ATTCGCCCGG GCCCTGCCCG 7200 TCTGGGCGCAA ATTCTGCGGA AGTCTCGGAG ATTCGCCCGG GCCCTGCCCG 7300 GTCCCCTCCT GTGCCTCCGC CTCGGAAAAA GCGTACGGTG GTCCTCACCG 7350 AATCAACCCT ATCTACTGC TTGGCCGAGC TTGCCACCAA AAGTTTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGGCGAC ATTCCCCCAA AAGTTTTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGGCGAC ATTCCCCCAAA AAGTTTTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGGCGAC AATAACAAA CATCCTCTGA 7450 GCCCGCCCCT TCTGGCCACT CCCCCCGACTC CAACGTTGAG TCCTATTCTT 7500 CCATGCCCCCC CCTGCAACGG GAGCCTGGGG ATCCCCCACG 7350	CCIGGGATIC	CCITIGIGIC	CIGCCAGCGC	GGGTATAGGG	GGGICIGGGG	6400
AACATGTGCA GTGGCACGTT CCCCATTAAC GCCTACACCA CGGCCCCTG 6550 TACTCCCCTT CCTGGGCCA ACTATAAGTT CGGCTGTGG AGGTGTCTG 6600 CACAGCAATA CGTGCACATA AGGCGGTGG GGCACTTCCA CTACGTATCG 6650 GGTATCACTA CTGACAATCT TAAATGCCGG TGCCACATCC CATCGCCCA 6700 ATTITTCACA GAATTGCACG GGGTGCGCCT ACACAGGTTT GCGCCCCTT 6750 GCAAGCCCTT GCTGCGCCAG GAGGTATCAT TCACAGTATGG ACTCCACGAG 6800 TACCCGGTGG GGTCCCAATT ACCTTGCCAG CCCGAACCGG ACGTAGCCGT 6850 GTTGACGTCC ATCCTCACTG ATCCTCCCA TATAACAGCA CAGGGGGCG 6900 GGAGAAGGTT GCCGACAGGG TCACCCCCTT CTATGGCCAG CTCCTCGGCT 6950 AGCCAGCTGT CCGCTCCATC TCTCAAGGCA ACTTGCACGG CCAACCATGA 7000 CTCCCCTGAC GCCGACCTCA TACAGGCTAA CCTCCTGTGG AGGCAGCACA 7050 TGGGCGCCAA CATCACCAGG GTTGAGTCAG ACAACAAAGT GGTGATTCTG 7100 GACTCCTTCG ATCCCTTTGT GCCACAGCGA GATCACCGG CCCTGCCCG 7200 TCTGGGCGCG GCCGGACTTCA AACCCCCCCC TACTACCAG GCCCTGCCCG 7200 TCTGGGCGCG GCCGGACTTCA AACCCCCCCC TACTACACAC GCCCTGCCCG 7200 TCTGGGCGCG GCCGGACTTC AACCCCCCCC TACTACACAC GTGCAAAAAAG 7250 CCTGACTACG AACCACCTGT GGTCCATGCC TCCCCCCTTAC CACCTCCACG 7300 GTCCCCTCCT GTGCCTCCCC CTCGGAAAAAA GCGTTACGGTG GTCCTCACCG 7350 AATCAACCCT ATCTGCGCAC TTGGCCCACCAA AAGTTTTGCC 7400 AGCTCCTCAA CTTCCGGCCAT TACGGGCGAC CACCGCACAC CACCTCCACG 7350 CCATGCCCCCC TCTGGGCGCC CCCCCGACTC CGACGTTCAG TCCTATTCTT 7500 CCATGCCCCC CCTGGAGGGG GAGCCTGCCC AATCCCCCCC TCCTATTCTTT 7500 CCATGCCCCCC TCTGGGCGCC CCCCCGACTC CGACGTTCAG TCCTATTCTTT 7500	AGGAGACGGC	ATTATGCACA	CICGCIGCCA	CIGIGGAGCI	GAGATCACTG	6900
TACTOCCCTT CCTGCGCCGA ACTATAAGIT CCCCCTGTGG AGGGTGTCTG 6600 CAGAGGAATA CGTGGAGATA AGGCGGGTGG GGGACTTCCA CTACGTATCG 6650 GGTATGACTA CTGACAATCT TAAATGCCG TGCCAGATCC CATCGCCGA 6700 ATTITITCACA GAATTGGACG GGGTGCGCCT ACACAGGTTT GCGCCCCTT 6750 GCAAGCCTT GCTGCGGGGG GAGGTATCAT TCAGAGTAGG ACTCCACGAG 6800 TACCCGGTGG GGTGCAATT ACCTTGCGAG CCCCAACCGG ACGTAGCCGT 6850 GTTGACGTCC ATGCTCACTG ATCCCTCCCA TATAACAGCA GAGGGGGCG 6900 GGAGAAGGTT GGCGACAGG TCACCCCTT CTATGGCCAG CTCCTGCGCT 6950 AGCCAGCTGT CCGCTCCATC TCTCAAGGCA ACTTGCACCG CTCCTGCGCT 6950 CTCCCCTGAC GCCGACCTCA TAGAGGCTAA CCTCCTGTGG AGGCAGGACA 7000 CTCCCCTGAC GCCGACCTCA TAGAGGCTAA CCTCCTGTGG AGGCAGGACA 7050 TGGGCGGCAA CATCACCAGG GTTGAGTCAG ACAACAAAGT GGTGATTCTG 7100 GACTCCTTCG ATCCCCTTGT GGCAGAGGAG GATCACCGG ACGTCCCCG 7200 TCTGGGCGGAA ATTCTGCGGA AGTCTCGCAG ATTCGCCCGG GCCCTGCCCG 7200 TCTGGGCGGAA ATTCTGCGGA AGTCTCGCAG ATTCGCCCGG GCCCTCCCCG 7300 GTCCCCTCCT GTGCCTCGC CTCGGAAAAA GCGTACGGTG GTCCTCACG 7350 AATCAACCCT ATCTACTGCC TTGGCCGAGC TTGCCCACAA AAGTTTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGGCGAC TTGCCCACAA AAGTTTTGGC 7400 AGCTCCTCAA CTTCCGCCAT TACGGGCGAC TTGCCCACAA AAGTTTTGGC 7450 GCCCCCCCCT TCTGGCCTGCC CCCCCGACTC CGACGTTGAG TCCTATTCTT 7500 CCATGCCCCC CCTGCAGGGG GAGCCTGGGG ATCCCGATCT CAGCGACGGG 77550	GACATGICAA	AAACGGGACG	ATGAGGATCG	TCGGTCCTAG	GACCTGCAGG	6950
CAGAGGAATA CGTGGAGATA AGGCGGGTGG GGGACTTCCA CTACGTATICG 6650 GGTATGACTA CTGACAATCT TAAATGCCG TGCCAGATCC CATCGCCGA 6700 ATTITTCACA GAATTGGACG GGGTGCGCCT ACACAGGTTT GCGCCCCCTT 6750 GCAAGCCCTT GCTGCGGGAG GAGGTATCAT TCAGAGTAGG ACTCCACGAG 6800 TACCCGGTGG GGTCGCAATT ACCTTGCGAG CCCGAACCGG ACGTAGCCGT 6850 GTTGACGTCC ATGCTCACTG ATCCCTCCCA TATAACAGCA GAGGCGCCG 6900 GGAGAAGGTT GGCGAGAGG TCACCCCTT CTATGGCCAG CTCCTCGGCT 6950 AGCCAGCTGT CCGCTCCATC TCTCAAGGCA ACTTGCACCG CCAACCATGA 7000 CTCCCCTGAC GCCGAGCTCA TAGAGGCTAA CCTCCTGTGG AGGCAGGAGA 7050 TGGGCGGCAA CATCACCAGG GTTGAGTCAG AGAACAAAGT GGTGATTCTG 7100 GACTCCTTCG ATCCGCTTGT GGCAGAGGAG GATGAGCGG ACGTCCGGT 7150 ACCTGCAGAA ATTCTGCGGA AGTCTCGGAG ATTCGCCCG GCCCTGCCCG 7200 TCTGGCCGCG GCCGGACTAC AACCCCCGC TAGTAGAGAC GTGGAAAAAAG 7250 CCTGACTACG AACCACCTGT GGTCCATGGC TGCCCGCTAC CACCTCCACG 7300 GTCCCCTCCT GTGCCTCGC CTCGCAAAAAA GCGTACGGTG GTCCTCACCG 7350 AATCAACCCT ATCTACTGCC TTGGCCGAGC TTGCCACCAA AAGTTTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGGCGAC AATACGACAA AAGTTTTGGC 7450 GCCCGCCCCT TCTGGCCACT TACGGCCGC CACCTCCTCA 7450 GCCCGCCCCT TCTGGCCACT TACGGCCGC CACCTCTCTCA 7450 GCCCGCCCCT TCTGGCCACT CCCCCCGCCCC CACCTCCTCT 7500 CCATGCCCCC CCTGGAGGGG GAGCCTGGGG ATCCGGATCT CAGCCACCG 7550	AACATGTGGA	GTGGGACGTT	CCCCATTAAC	GCCTACACCA	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	6550
GETATGACTA CTGACAATCT TAAATGCCCG TGCCAGATCC CATGGCCCGA 6700 ATTITITCACA GAATTGGACG GEGTGCGCCT ACACAGGTTT GCGCCCCCTT 6750 GCAAGCCCTT GCTGCGGGAG GAGGTATCAT TCAGAGTAGG ACTCCACGAG 6800 TACCCGGTGG GGTCGCAATT ACCTTGCGAG CCCGAACCGG ACGTAGCGGT 6850 GTTGACGTCC ATGCTCACTG ATCCCTCCCA TATAACAGCA GAGGGGCCG 6900 GGAGAAGGTT GGCGAGAGG TCACCCCCTT CTATGGCCAG CTCCTCGGCT 6950 AGCCAGCTGT CCGCTCCATC TCTCAAGGCA ACTTGCACCG CCAACCATGA 7000 CTCCCCTGAC GCCGAGCTCA TAGAGGCTAA CCTCCTGTGG AGGCAGGAGA 7050 TGGGCGCCAA CATCACCAGG GTTGAGTCAG AGAACAAAGT GGTGATTCTG 7100 GACTCCTTCG ATCCCCTTGT GGCAGAGGAG GATGAGCGG ACGTCCCGT 7150 ACCTGCAGAA ATTCTGCGGA AGTCTCGGAG ATTCGCCCG GCCCTGCCCG 7200 TCTGGCCGCG GCCGGACTAC AACCCCCGC TAGTAGAGAC GTGCAAAAAG 7250 CCTGACTACG AACCACCTGT GGTCCATGGC TGCCCGCTAC CACCTCCACG 7300 GTCCCCTCCT GTGCCTCGC CTCGCAAAAAA GCGTACGGTG GTCCTCACCG 7350 AATCAACCCT ATCTACTGCC TTGCCCGAGC TTGCCACCAA AAGTTTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGCCGAC TTGCCACCAA AAGTTTTGGC 7450 GCCCGCCCCT TCTGGCCACT TACGGCCGCAC CACCTCCTCA 7450 GCCCGCCCCT TCTGGCCACT TACGGCCGC CCACCTTCACCG 7550 CCATGCCCCC CCTGGAGGGG GAGCCTGGGG ATCCGCATCT CAGCCACCAA TAGTTTTCTT 7500 CCATGCCCCC CCTGGAGGGG GAGCCTGGGG ATCCGCATCT CAGCCACCGG 7550	TACTCCCCTT	CCTGCGCCGA	ACTATAAGIT	CCCCCIGICG	AGGGIGICIG	6600
ATTITICACA GAATIGGACG GGGIGGGCCT ACACAGGITT GCGCCCCCTT 6750 GCAAGCCCTT GCIGGGGAG GAGGIAICAT TCAGAGIAGG ACTCCACGAG 6800 TACCCGGIGG GGIGGCAATT ACCTTGCGAG CCCGAACGG ACGIAGCGGT 6850 GITGACGICC ATGCTCACTG ATCCCTCCCA TATAACAGCA GAGGGGGCG 6900 GGAGAAGGIT GGCGAGAGGG TCACCCCCTT CIATGGCCAG CTCCTCGGCT 6950 AGCCAGCTGT CCGCTCCATC TCTCAAGGCA ACTTGCACCG CCAACCATGA 7000 CTCCCCTGAC GCCGAGCTCA TAGAGGCTAA CCTCCTGIGG AGGCAGGAGA 7050 TGGGCGGCAA CATCACCAGG GITGAGICAG AGAACAAAGT GGIGATTCTG 7100 GACTCCTTCG ATCCGCTTGT GGCAGAGGAG GATGAGCGG AGGICTCCGT 7150 ACCTGCAGAA ATTCTGCGGA AGICTCGGAG ATTCGCCCGG GCCCTGCCCG 7200 TCTGGGCGGG GCCGGACTTAC AACCCCCCGC TAGTAGAGCAC GIGGAAAAAG 7250 CCTGACTACG AACCACCTGT GGTCCATGGC TGCCCGCTTAC CACCTCCACG 7300 GTCCCCTCCT GTGCCTCCGC CTCGGAAAAAA GCGTACCGTG GTCCTCACCG 7350 AATCAACCCT ATCTACTGCC TTGGCCGAGC TTGCCCACAA AAGITTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGGCGAC AATACGACAA CATCCTCTGA 7450 GCCCGCCCCT TCTGGCTGCC CCCCCCACTC CGACGTTCAG TCCTATTCTT 7500 CCATGCCCCCC CCTGGAGGGG GAGCCTGGGG ATCCCCACG TCCTATTCTT 7500	CAGAGGAATA	CGTGGAGATA	AGGCGGGTGG	GGGACTICCA	CTACGTATCG	6650
GCAAGCCTT GCTGCGGGAG GAGGTATCAT TCAGAGTAGG ACTCCACGAG 6800 TACCCGGTGG GGTCGCAATT ACCTTGCAG CCCGAACCG ACGTAGCCGT 6850 GTTGACGTCC ATCCTCACTG ATCCCTCCCA TATAACAGCA GAGGGGGCCG 6900 GGAGAAGGTT GCCGAGAGG TCACCCCTT CTATGGCCAG CTCCTCGGCT 6950 AGCCAGCTGT CCGCTCCATC TCTCAAGGCA ACTTGCACCG CCAACCATGA 7000 CTCCCCTGAC GCCGAGCTCA TAGAGGCTAA CCTCCTGTGG AGCCAGGAGA 7050 TGGGCGGCAA CATCACCAGG GTTGAGTCAG AGAACAAAGT GGTGATTCTG 7100 GACTCCTTCG ATCCGCTTGT GGCAGAGGAG GATCACCGG AGGTCTCCGT 7150 ACCTGCAGAA ATTCTGCGGA AGTCTCGGAG ATTCGCCCGG GCCCTGCCG 7200 TCTGGGGGGG GCCGGACTAC AACCCCCGC TAGTAGAGAC GTGGAAAAAG 7250 CCTGACTACG AACCACCTGT GGTCCATGGC TGCCCGCTAC CACCTCCACG 7300 GTCCCCTCCT GTGCCTCCGC CTCGGAAAAAA GCGTACGGTG GTCCTCACCG 7350 AATCAACCCT ATCTACTGCC TTGGCCGAGC TTGCCCACCAA AAGTTTTGGC 7400 ACCTCCTCAA CTTCCGGCAT TACGGGCGAC AATACGACAA CATCCTCTGA 7450 GCCCGCCCCT TCTGGCTGCC CCCCCGACTC CGACGTTGAG TCCTATTCTT 7500 CCATGCCCCC CCTGGGAGGGG GAGCCTGGGG ATCCGGATCT CAGCGACGGG 7550	GCTATGACTA	CTGACAATCT	TAAATGCCCG	TECCAGATCC	CATOGCCCCA	6700
TACCOGGIGG GGICGCAATT ACCTIGGGAG CCCGAACCGG ACGIAGCCGT 6850 GTIGACGICC ATGCICACTG ATCCCICCCA TATAACAGCA GAGGGGCCG 6900 GGAGAAGGIT GGCGAGAGGG TCACCCCCTT CTATGGCCAG CTCCTCGGCT 6950 AGCCAGCIGT CCGCICCATC TCTCAAGGCA ACTTGCACCG CCAACCATGA 7000 CTCCCCTGAC GCCGAGCICA TAGAGGCTAA CCTCCTGIGG AGGCAGGAGA 7050 TGGGCGGCAA CATCACCAGG GTTGAGTCAG ACAACAAAGT GGIGATTCTG 7100 GACTCCTTCG ATCCGCTTGT GGCAGAGGAG GATGAGCGG ACGTCCCGT 7150 ACCTGCAGAA ATTCTGCGGA AGTCTCGGAG ATTCGCCCG GCCCTGCCCG 7200 TCTGGGCGG GCCGGACTAC AACCCCCGC TAGTAGAGAC GIGGAAAAAG 7250 CCTGACTACG AACCACCTGT GGTCCATGGC TGCCCGCTAC CACCTCCACG 7300 GTCCCCTCCT GTGCCTCCGC CTCGGAAAAA GCGTACGGTG GTCCTCACCG 7350 AATCAACCCT ATCTACTGCC TTGGCCGACC TTGCCCACCAA AAGTTTTGGC 7400 AGCTCCTCAA CTTCCGCCAT TACGGGCGAC AATACCACAA CATCCTCTGA 7450 GCCCGCCCCT TCTGGCTGCC CCCCCGACTC CGACGTTGAG TCCTATTCTT 7500 CCATGCCCCC CCTGGGAGGGG GAGCCTGGGG ATCCGCATCT CAGCGACGGG 7550	ATTTTTCACA	GAATTGGACG	GGGTGCGCCT	ACACAGGITT	GCGCCCCTT	6750
GITGACGICC ATGCTCACTG ATCCCICCCA TATAACAGCA GAGGGGGCG 6900 GGAGAAGGIT GGCGAGAGG TCACCCCCTT CTATGGCCAG CTCCTGGCT 6950 AGCCAGCIGT CCGCTCCATC TCTCAAGGCA ACTTGCACCG CCAACCATGA 7000 CTCCCCTGAC GCCGAGCTCA TAGAGGCTAA CCTCCTGIGG AGGCAGGAGA 7050 TGGGCGCAA CATCACCAGG GITGAGICAG AGAACAAAGT GGIGATTCTG 7100 GACTCCTTCG ATCCGCTTGT GGCAGAGGAG GATGAGCGGG AGGICTCCGT 7150 ACCTGCAGAA ATTCTGCGGA AGTCTCGCAG ATTCGCCCGG GCCCTGCCCG 7200 TCTGGGCGGG GCCGGACTAC AACCCCCCGC TAGTAGAGAC GTGGAAAAAG 7250 CCTGACTACG AACCACCTGT GGICCATGGC TGCCCGCTAC CACCTCCACG 7300 GTCCCCTCCT GTGCCTCCGC CTCGGAAAAA GCGTACGGTG GTCCTCACCG 7350 AATCAACCCT ATCTACTGCC TTGGCCGAGC TTGCCCACCAA AAGTTTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGGCGAC AATACCACAA CATCCTCTGA 7450 GCCCGCCCCT TCTGGCTGCC CCCCCGACTC CGACGTTGAG TCCTATTCTT 7500 CCATGCCCCC CCTGGAGGGG GAGCCTGCGG ATCCGCATCT CAGCGACGGG 7550	GCAAGCCCTT	GCTGCGGGAG	GAGGIATCAT	TCAGAGTAGG	ACTOCACGAG	6800
GGAGAAGGIT GGCGAGAGG TCACCCCCTT CTATGGCCAG CTCCTCGGCT 6950 AGCCAGCTGT CCGCTCCATC TCTCAAGGCA ACTTGCACCG CCAACCATGA 7000 CTCCCCTGAC GCCGAGCTCA TAGAGGCTAA CCTCCTGTGG AGGCAGGAGA 7050 TGGGCGGCAA CATCACCAGG GTTGAGTCAG ACAACAAAGT GGTGATTCTG 7100 GACTCCTTCG ATCCGCTTGT GGCAGAGGAG GATGAGCGGG AGGTCTCCGT 7150 ACCTGCAGAA ATTCTGCGGA AGTCTCGGAG ATTCGCCCGG GCCTGCCCG 7200 TCTGGGCGGG GCCGGACTAC AACCCCCGC TAGTAGAGAC GTGCAAAAAG 7250 CCTGACTACG AACCACCTGT GGTCCATGGC TGCCCGCTAC CACCTCCACG 7300 GTCCCCTCCT GTGCCTCCGC CTCGGAAAAA GCGTACGGTG GTCCTCACCG 7350 AATCAACCCT ATCTACTGCC TTGGCCGACC TTGCCCACCAA AAGTTTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGCCGAC AATACGACAA CATCCTCTGA 7450 GCCCGCCCCT TCTGGCTGCC CCCCCGACTC CGACGTTGAG TCCTATTCTT 7500 CCATGCCCCC CCTGGAGGGG GAGCCTGGGG ATCCGGATCT CAGCGACGGG 7550	TACCCGGTGG	GGTCGCAATT	ACCTTGCGAG	CCCGAACCGG	ACGIAGCCGI	6850
AGCCAGCIGT CCGCTCCATC TCTCAAGGCA ACTTGCACCG CCAACCATGA 7000 CTCCCCTGAC GCCGAGCTCA TAGAGGCTAA CCTCCTGTGG AGGCAGGAGA 7050 TGGGCGGCAA CATCACCAGG GTTGAGTCAG AGAACAAAGT GGTGATTCTG 7100 GACTCCTTCG ATCCGCTTGT GGCAGAGGAG GATGAGGGG AGGTCTCCGT 7150 ACCTGCAGAA ATTCTGCGGA AGTCTCGGAG ATTCGCCCGG GCCCTGCCCG 7200 TCTGGGCGCG GCCGGACTAC AACCCCCCGC TAGTAGAGAC GTGGAAAAAAG 7250 CCTGACTACG AACCACCTGT GGTCCATGGC TGCCCGCTAC CACCTCCACG 7300 GTCCCCTCCT GTGCCTCCGC CTCGGAAAAA GCGTACGGTG GTCCTCACCG 7350 AATCAACCCT ATCTACTGCC TTGGCCGAGC TTGCCCACCAA AAGTTTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGGCGAC AATACGACAA CATCCTCTGA 7450 GCCCGCCCCT TCTGGCTGCC CCCCCGACTC CGACGTTGAG TCCTATTCTT 7500 CCATGCCCCC CCTGGAGGGG GAGCCTGGGG ATCCGCATCT CAGCGACGGG 7550	GITGACGICC	ATGCTCACTG	ATCCCTCCCA	TATAACAGCA	CAGCCCCC	6900
CTCCCCTGAC GCCGAGCTCA TAGAGGCTAA CCTCCTGTGG AGGCAGGAGA 7050 TGGGCGGCAA CATCACCAGG GTTGAGTCAG AGAACAAAGT GGTGATTCTG 7100 GACTCCTTCG ATCCGCTTGT GGCAGAGGAG GATGAGCGGG AGGTCTCCGT 7150 ACCTGCAGAA ATTCTGCGGA AGTCTCGGAG ATTCGCCCGG GCCCTGCCCG 7200 TCTGGGCGCG GCCGGACTAC AACCCCCGC TAGTAGAGAC GTGGAAAAAG 7250 CCTGACTACG AACCACCTGT GGTCCATGGC TGCCCGCTAC CACCTCCACG 7300 GTCCCCTCCT GTGCCTCCGC CTCGGAAAAA GCGTACGGTG GTCCTCACCG 7350 AATCAACCCT ATCTACTGCC TTGGCCGAGC TTGCCCACCAA AAGTTTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGGCGAC AATACGACAA CATCCTCTGA 7450 GCCCGCCCCT TCTGGCTGCC CCCCCGACTC CGACGTTGAG TCCTATTCTT 7500 CCATGCCCCC CCTGGAGGGG GAGCCTGGGG ATCCGGATCT CAGCGACGGG 7550	GGAGAAGGTT	CCCCACAGGG	TCACCCCCTT	CTATGGCCAG	CICCICGGCT	6950
TOGOCOGOCA CATCACCAGO GITGAGICAG AGAACAAAGT GGIGATICTG 7100 GACTCCTTCG ATCCGCTTGT GGCAGAGGAG GATGAGCGG AGGTCTCCGT 7150 ACCTGCAGAA ATTCTGCGGA AGTCTCGGAG ATTCGCCCGG GCCCTGCCCG 7200 TCTGGGCGGG GCCGGACTAC AACCCCCCGC TAGTAGAGAC GIGGAAAAAG 7250 CCTGACTACG AACCACCTGT GGTCCATGGC TGCCCGCTAC CACCTCCACG 7300 GICCCCTCCT GTGCCTCGC CTCGGAAAAA GCGTACGGTG GTCCTCACCG 7350 AATCAACCCT ATCTACTGCC TTGGCCGAGC TTGCCCACCAA AAGTTTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGGCGAC AATACGACAA CATCCTCTGA 7450 GCCCGCCCCT TCTGGCTGCC CCCCCGACTC CGACGTTGAG TCCTATTCTT 7500 CCATGCCCCC CCTGGAGGGG GACCCTGGGG ATCCGGATCT CAGCGACGGG 7550	AGCCAGCIGI	CCGCTCCATC	TCTCAAGGCA	ACTIGCACCG	CCAACCAIGA	7000
CACTOCTICG ATCOCCTICT GCCAGAGGAG GATGAGGGG AGGICTCCGT 7150 ACCTGCAGAA ATTCTGCGGA AGTCTCGGAG ATTCGCCCGG GCCCTGCCCG 7200 TCTGGGCGG GCCGGACTAC AACCCCCGC TAGTAGAGAC GTGGAAAAAG 7250 CCTGACTACG AACCACCTGT GGTCCATGGC TGCCCGCTAC CACCTCCACG 7300 GTCCCCTCCT GTGCCTCCGC CTCGGAAAAA GCGTACGGTG GTCCTCACCG 7350 AATCAACCCT ATCTACTGCC TTGGCCGAGC TTGCCCACCAA AAGTTTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGGCCAC AATACGACAA CATCCTCTGA 7450 GCCCGCCCCT TCTGGCTGCC CCCCCGACTC CGACGTTGAG TCCTATTCTT 7500 CCATGCCCCC CCTGGAGGGG GAGCCTGGGG ATCCGGATCT CAGCGACGGG 7550						7050
ACCTGCAGAA ATTCTGCGGA AGTCTCGGAG ATTCGCCCGG GCCCTGCCCGG 7200 TCTGGGCGCG GCCGGACTAC AACCCCCCGC TAGTAGAGAC GIGGAAAAAG 7250 CCTGACTACG AACCACCTGT GGTCCATGGC TGCCCGCTAC CACCTCCACG 7300 GTCCCCTCCT GTGCCTCCGC CTCGGAAAAA GCGTACGGTG GTCCTCACCG 7350 AATCAACCCT ATCTACTGCC TTGGCCGAGC TTGCCACCAA AAGTTTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGGCGAC AATACGACAA CATCCTCTGA 7450 GCCCGCCCCT TCTGGCTGCC CCCCCGACTC CGACGTTGAG TCCTATTCTT 7500 CCATGCCCCC CCTGGAGGGG GAGCCTGGGG ATCCGGATCT CAGCGACGGG 7550	TGGGCGGCAA	CATCACCAGG	GTTGAGTCAG	AGAACAAAGT	GGIGATICIG	7100
TCTGGGCGG GCCGGACTAC AACCCCCGC TAGTAGAGAC GTGGAAAAAG 7250 CCTGACTACG AACCACCTGT GGTCCATGGC TGCCGCCTAC CACCTCCACG 7300 GTCCCCTCCT GTGCCTCCGC CTCGGAAAAA GCGTACGGTG GTCCTCACCG 7350 AATCAACCCT ATCTACTGCC TTGGCCGAGC TTGCCACCAA AAGTTTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGGCGAC AATACGACAA CATCCTCTGA 7450 GCCCGCCCCT TCTGGCTGCC CCCCCGACTC CGACGTTGAG TCCTATTCTT 7500 CCATGCCCCC CCTGGAGGGG GAGCCTGGGG ATCCGGATCT CAGCGACGGG 7550	GACTCCTTCG	ATCCCCTTCT	GGCAGAGGAG	CATCACCGGG	AGGICICCGT	
CCTGACTACG AACCACCTGT GGTCCATGGC TGCCGCTAC CACCTCCACG 7300 GTCCCCTCCT GTGCCTCCGC CTCGGAAAAA GCGTACGGTG GTCCTCACCG 7350 AATCAACCCT ATCTACTGCC TTGGCCGAGC TTGCCACCAA AAGTTTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGGCGAC AATACGACAA CATCCTCTGA 7450 GCCCGCCCCT TCTGGCTGCC CCCCCGACTC CGACGTTCAG TCCTATTCTT 7500 CCATGCCCCC CCTGGAGGGG GAGCCTGGGG ATCCGGATCT CAGCGACGGG 7550	ACCIGCAGAA	ATTCTGCGGA	AGICICGGAG	ATTOGCCCGG	GCCIGCCC	
GICCCCTCCT GIGCCTCCGC CTCGGAAAAA GCGTACGGIG GICCTCACCG 7350 AATCAACCCT ATCTACTGCC TTGGCCGAGC TTGCCACCAA AAGITTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGGCGAC AATACGACAA CATCCTCTGA 7450 GCCCGCCCCT TCTGGCTGCC CCCCCGACTC CGACGTTGAG TCCTATTCTT 7500 CCATGCCCCC CCTGGAGGGG GAGCCTGGGG ATCCGGATCT CAGCGACGGG 7550	TCTGGGGGG	GCCGGACIAC	AACCCCCCCCC	TAGTAGAGAC	GIGGAAAAAG	
AATCAACCCT ATCTACTGCC TTGGCCGAGC TTGCCACCAA AAGITTTGGC 7400 AGCTCCTCAA CTTCCGGCAT TACGGGCGAC AATACGACAA CATCCTCTGA 7450 GCCCGCCCCT TCTGGCTGCC CCCCCGACTC CGACGTTGAG TCCTATTCTT 7500 CCATGCCCCC CCTGGAGGGG GAGCCTGGGG ATCCGGATCT CAGCGACGGG 7550	CCIGACIACG	AACCACCIGI	GGTCCATGGC	TGCCCGCTAC	CACCTCCACG	
AGCICCICAA CITCOGGCAT TACGGGGGAC AATACGACAA CATCCICIGA 7450 GCCCGCCCCT TCTGGCTGCC CCCCCGACTC CGACGITGAG TCCTATTCTT 7500 CCATGCCCCC CCTGGAGGGG GAGCCTGGGG ATCCGGATCT CAGCGACGGG 7550						
GCCCGCCCCT TCTGGCTGCC CCCCCGACTC CGACGTTGAG TCCTATTCTT 7500 CCATGCCCCC CCTGGAGGGG GAGCCTGGGG ATCCGGATCT CAGCGACGGG 7550						
CCATGCCCC CCTGGAGGG GAGCCTGGGG ATCCGGATCT CAGCGACGGG 7550						
TCATGGICGA CGGICAGIAG TGGGGCCCAC ACGGAAGATG TCGIGIGCIG 7600					•	
	TCATGGTCGA	CGGICAGIAG	TGGGGGGGAC	ACGGAAGATG	TEGIGICEIG	7600

FIG. 4D

10 20	30	40	50	
1234567890 1234567890 1	<u> 1234567890</u>	1234567890	<u>1234567890</u>	
CICAATGICT TATTCCIGGA	CAGGCGCACT	CGICACCCCG	TECECTECES	7650
AAGAACAAAA ACTGCCCATC				7700
CACAATCIGG TGIATICCAC				7750
CAAAGICACA TITICACACAC	IGCAAGIICI	GGACAGCCAT	TACCAGGACG	7800
TOCTCAAGGA GOTCAAAGCA (ECCECCICAA	AAGIGAAGGC	TAACTIGCIA	7850
TOOGRAGAGG AAGCTTGCAG				7900
GITTGCCIAT GGGGCAAAAG	ACCICCCITG	CCATCCCAGA	AAGGCCGIAG	7950
CCCACATCAA CTCCGTGTGG	AAACACCTIC	TGGAAGACAG	TGIAACACCA	8000
ATAGACACTA CCATCATGGC	CAAGAACGAG	GTTTTCTGCG	TICAGCCIGA	8050
CAAGGGGGT CGTAAGCCAG	CICCICAT	CGIGITCCCC	GACCIGGGCG	8100
TECCCCTCTC CCACAACATC	GCCCIGIACG	ACGIGGITAG	CAAGCTCCCC	8150
CIGGCCGIGA TGGGAAGCIC	CTACGGATTC	CAATACTCAC	CAGGACAGCG	8200
GGITGAATTC CTCGTGCAAG	CGIGGAAGIC	CAAGAAGACC	CCGATGGGGT	8250
TCTCGTATGA TACCCGCTGT				8300
CGTACGGAGG AGGCAATTTA	CCAATGITGI	GACCIGGACC	CCCAAGCCCG	8350
CGIGGCCATC AAGICCCICA	CIGAGAGGCI	TIATGITGGG	GGCCCICTIA	8400
CCAATICAAG GGGGAAAAC	TGCGGCTACC	CAGGIGCCG	CGCGAGCGGC	8450
GIACIGACAA CIAGCIGIGG	TAACACCCIC	ACTIGCTACA	TCAAGGCCCG	8500
GCAGCCIGT CGAGCCGCAG	GGCTCCAGG2	L CIGCACCAIC	CICGIGIGIG	8550
GCCACCACIT AGICGITATC	TGTGAAAGIC	CGGGGGTCCA	. GGAGGACGCG	8600
COCACCCICA CACCCTICAC	GGAGGCIATO	ACCAGGIACI	, caecacacac	8650
CGGGGACCCC CCACAACCAG	AATACGACT	r ggagcitata	ACATCATECT	8700
CCTCCAACGT GTCAGTCGCC	CACGACGGO	G CTGGAAAGAC	GGICIACIAC	8750
			GGGAGACAGC	8800
AAGACACACT CCAGTCAATT	CCTGGCTAG	G CAACATAATO	AIGITIGCCC	8850
CCACACIGIG GGCGAGGAIG	ATACIGATG	A COCATTICIT	TAGCGICCIC	8900
ATACTCACCO ATCACCITGA	ACAGGCTCT	r aacigigag	A TCTACGGAGC	8950
CICCIACICC ATAGAACCAC	TGGATCIAO	C TOCAATCAT	r caaagactcc	9000
ATGGCCTCAG CGCATTITICA	CICCACAGI	T ACICICCAG	G TGAAATCAAT	9050
AGGGTGGCCG CATGCCTCAG	AAAACTTGG	G GICCCGCCC	r teceaectie	9100
GAGACACCGG GCCCGGAGCG	TCCGCGCTA	G GCTTCTGTC	C AGAGGAGGCA	9150
GGGCTGCCAT ATGTGGCAAG	TACCICITO	'A ACTGGGCAG	T AAGAACAAAG	9200
CTCAAACTCA CTCCAATAGO	GGCCGCTGC	CCCCCCCAC	T TGICCGGIIG	9250
GITCACGGCT GGCTACAGCC	GGGGAGAC	IT TTATCACAC	C GIGICICAIG	9300
CCCCCCCCC CTCCTTCTCC	TTTTGCCT	AC TOCTOCTO	C TGCAGGGGIA	9350
GGCATCTACC TOCTCOCCA	A CCGATGAAC	G TIGGGGIAA	A CACTOOGGCC	9400
TCTTAACCCA TITCCTGTT	r TTTTTTTT	er trrtrrrr	T TTTTCTTT	9450
TITTTTCTT TCCTTCCT	r Crittiri	C TITCITI	C CCTTCTTIAA	9500

FIG. 4E

H77C

10	20	30	40	50	,
1234567890	1234567890	1234567890	1234567890	1234567890	
TEGTESCICC	ATCTTAGCCC	TAGTCACGGC	TAGCIGIGAA	AGGICCGIGA	9550
GCCGCATGAC	TGCAGAGAGT	GCTGATACTG	GCCICICIGC	AGATCATGT	9599

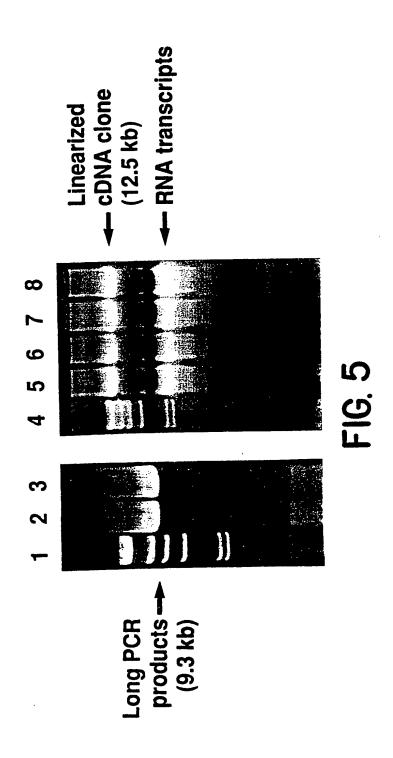
FIG. 4F

10 20 30 40 50	
1234567890 1234567890 1234567890 1234567890 1234567890	
MSTNEKBORK TKRNINRREQ DVKFEGGGQI VGGVYLLERR GERLGVRATR	50
KTSERSOPRG RROPIPKARR PECRIWAQEG YFWPLYGNEG CGWAGWLLSP	100
RGSRPSWGPT DPRRRSRNLG KVIDILITCGF ADLMGYIPLV GAPLGGAARA	150
LAHGYRVLED GYNYATGNLP GCSFSIFLLA LLSCLIVPAS AYOVRNSSGL	200
VHUTNDCENS SIVYEAADAI LHIPGCVPCV REGNASROW AVIPIVATRD	250
CKLPTTOLER HIDLLYGSAT LCSALYVGDL CGSVFLVGQL FIFSPRRHWT	300
TOTONISTYP CHICHRAM DAMAWSPIA ALWAQLLRI PQAIMDMIAG	350
AHWOVIAGIA YESWOWAK VLVVILLEAG VDAEHHVIGG NAGRITAGIV	400
GLUTECAKON IOLININGSW HINSTALNON ESLNIGWLAG LFYCHKENSS	450
COPERLASOR RUIDEAOGNG PISYANGSGL DERPYOWHYP PRPOGIVPAK	500
SUCCEMPORT PSPVVGTID RSCAPTYSWG ANDIDVEVLN NIRPPLGNWF	550
CTWMNSTGF TKVCGAPPCV IGGVCNVTLL CPIDCFRKHP EATYSRCGSG	600
ENTTPROMID YPYRLWHYPC TINYTIFKVR MYVGGVEHRL EAACNWIRGE	650
RODIEDRORS ELSPLILSTY OWOVLPCSFT TLPALSTGLI HLHQNIVDVQ	700
VIVOVOSSTA SWATKWEYVV LLFLILADAR VCSCLWMLL ISQAFAALEN	750
INTINIASIA CIHCLUSELV FECEAWYLKG RWPGAVYAL YGWPLLLLL	800
LALBORAYAL DIEVAASOGG VVLVGLMALT LSPYYKRYIS WOMWLQYFL	850
TRYFAOIHW VPPINVRGGR DAVILIMOVV HPILVFDIIK LLLAIFGPLW	900
TIONSLIKUP YEVRVOGILR ICALARKIAG CHYVQMAIIK LGALIGIYVY	950
NHITTPLETMA HNGLEDLAVA VEPVVFSRME TKLITWGADT AACGDIINGL	1000
BYSARRODET LIGRADOMYS KOWRLLAPIT AYAQQIRGLL GCIITSLIGR	1050
DYNOVECENO TYSTATOTEL ATCINGVOWT VYHGAGIRII ASPKGPVIQM	1100
YTM/TODING WPAPOGSRSL TPCTCGSSDL YLVTRHADVI PVRRRGDSRG	1150
STISPRPTSY IKGSSGGPLL CPACHAVGLF RAAVCIRGVA KAVDFIPVEN	1200
LGTIMRSPVF TENSSPPAVP QSFQVAHLHA PIGSGKSIKV PAAYAAQGYK	1250
VLVLNPSVAA TLGFGAYMSK AHGVDPNIRT GVRTTTTGSP TTYSTYGKFL	1300
ADGGCSGGAY DILICDECHS TDATSILGIG TVLDQAETAG ARLVVLATAT	1350
PPGSVIVSHP NIEEVALSTT GEIPFYGKAI PLEVIKGGRH LIFCHSKKKC	1400
DELAAKLVAL GINAVAYYRG LDVSVIPISG DVVVVSIDAL MIGFIGDFDS	1450
VIDCNICVIQ TVDFSLDPIF TIETTILPQD AVSRIQRRGR TGRGKPGIYR	1500
EXABCERPS: MEDSVICEC YDAGCAWYEL TPAETIVRLR AYMNIEGLEV	1550
CODHLEFWEG VFIGLIHIDA HFLSQIKQSG ENFPYLVAYQ AIVCARAQAP	1600
PPSWDQMWKC LIRLKPTLHG PIPLLYRLGA VQNEVILTHP ITKYIMTOMS	1650
ADLEVVISIW VLVGGVLAAL AAYCLSIGCV VIVCRIVLSG KPAIIPDREV	1700 1750
LYOEFDEMEE CSOHLPYIEQ GMMLAEOFKO KALGLLQTAS RHAEVITPAV	1800
OTHORIES WAKHMANETS GIQYLAGIST LEGNEATASL MAFTAAVISE	1850
LTTGOTLLFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD	1900
ILAGYGAGVA GALVAFKIMS GEVPSTEDLV NLLPATLSPG ALVVGVVCAA	T200

FIG. 4G

10	20	30	40	50	
1234567890	1234567890	123456/890	1234567890	1234567890	1050
ILRRHVGPGE	GAVOWMVRLI	AFASRGNHVS	PIHYVPESDA	AARVIALLSS	1950
LIVIQLLRRL	HOWISSECTT	PCSGSWLRDI	WDWICEVLSD	FKIWLKAKLM	2000
POLPGIPFVS	CORGYRGWR	GDGIMHIRCH	CGAEITGHVK	NGIMRIVGPR	2050
			ALWRVSAEEY		2100
AASCALLINT	KCPCQIPSPE	FFTELDGVRL	HRFAPPCKPL	LREEVSFRVG	2150
LHEYPVGSQL	PCEPEPDVAV	LISMLIDPSH	TTAEAACRRL	ARGSPPSMAS	2200
SSASOLSAPS	LKATCIANHD	SPDAELIEAN	LLWRQEMGGN	ITRVESENKV	2250
VILDSFDPLV	AEEDEREVSV	PAEILRKSRR	FARALPWAR	PDYNPPLVET	2300
WKKPDYEPPV	VHGCPLPPPR	SPPVPPPRKK	RIVVLTESIL	STALAELATK	2350
SFGSSSTSGI	TGDNITISSE	PAPSGCPPDS	DVESYSSMPP	LEGEPGDPDL	2 40 0
SDGSWSTVSS	CADIEDWCC	SMSYSWIGAL	VIPCAAEEQK	LPINALSNSL	2450
LRHHNLVYST	TSRSACOROK	KVIFDRLQVL	DSHYQDVLKE	VKAAASKVKA	2500
NLLSVEEACS	LTPPHSAKSK	FGYGAKDVRC	HARKAVAHIN	SWKDLLEDS	2550
VTPIDITIMA	KNEVFCVOPE	KGGRKPARLI	VFPDLGVRVC	EKMALYDVVS	2600
KT.PT.AVMGSS	YGFOYSPGOR	VEFLVQAWKS	KKIPMGFSYD	TRCFDSIVIE	2650
SDIRTEFATY	CCDLDPOAR	VAIKSLITERL	YVGGPLINSR	GENCGYRRC R	2700
ASCAL TALESCE	MULTICYTKAR	AACRAAGLOD	CIMLVCGDDL	VVICESAGVQ	2750
			ELITSCSSNV		2800
			NIIMFAPILW		2850
יוובותסבוו	ONTNETYCA	CYSIEPLDLP	PIIQRLHGLS	AFSLHSYSPG	2900
	KI CAPPI RAW	RHRARSVRAR	LLSRGGRAAI	CCKYLFNWAV	2950
אבשבו הב נווטבא	AACRIDIGEN	FTAGYSGGDT	YHSVSHARPR	WFWFCILLLA	3000
AGVGTYLLPN					3011
PIOVOL X LUPIN	T.		A 1.1		

FIG. 4H



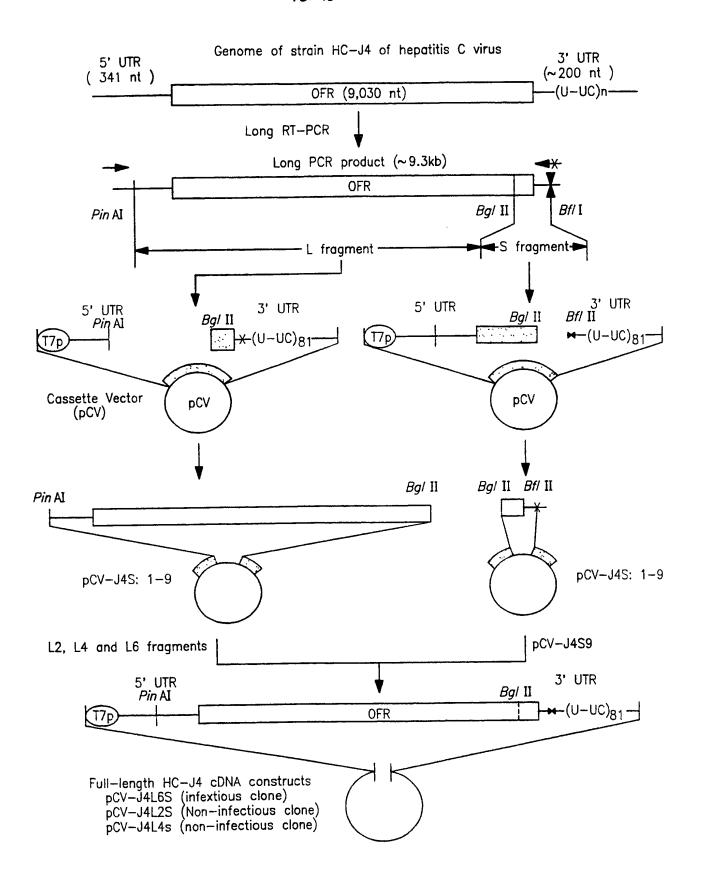


FIG. 6
SUBSTITUTE SHEET (RULE 26)

		Τ	_r		T	Т	_	_	7	_	Г	Т		Τ	Т	-	_	Т	7	
Cons-F	z			ΤΑ		K,U	¥	٥	2	∝	٥	٥	Z	2	2	Ŀ	ر	> -	₹	
Cons-D	٠		٠	-		K,0	۰			•		۰	•		•	•			•	
L4(C)			•	-		•	۰	,	,	۰		0		٢		•	,	,	٠	
L10(B)	,	,	•	1	-	0	۰	-	ı.	•		∢	ے	Š	٠	٠				
(1)*(B)		°	٠	F	-	o			•	۰		∢	ے	,	۰	٠		°	۰	
L3(B)		°	•	-	_	o	٥		I	ľ		∀		4	•	۰		٠	 	
(A)61		۰	۰		°	•	,	Ì	۰		•	•			•	A		•	•	
L8(A)		•	a		۰	0		°	•	C	3			·	٠				٠	
16(A)	<u> </u>	•			•	•		٥	•	ļ	-	۰		٥	٠			۰		
12(A)	2	•	·	ì	0	۰		۰	۰	ļ	2			٥	•	ŀ		>		,
(A)	2	v.			۰	۰		0	۰		۰	,		•	۰		4	۰		•
00-300	cd silos	Z	-	٦	⋖		٤.	∢	۵	2	∝	ر	٥	Z	z	-	الد	ں	-	1
framont	ר ונמלווופוור	18		36	52	25	2	682	305	25	231	077	557	234	250	200	588	304	OF F	- 5/c
			Core			-4														

FIG. 7A

15/49

													,					_					
Cons-F	E,T	H,Y	1,5	R,G	^	A,V	T	S	H,0	7,	A,T	S	S	A,V	٨	K,E	ľ,V	>	>	_		o	A
Cons-D	E,T	H,Y	1,5	G	•	۸	•	•	Ŧ		-	•	•	A,V	•	ш		•	•	•	•	•	•
L4(C)	•	•	•	•	•	^	œ	•			_	Р	•		•	Ш	I	•	•	•	•	•	•
L10(B)	-	۸	S	ပ	A	•	~	٠	エ		1	•	•	۸	•	Lυ		•	•	•	•	•	•
(8),		^	S	ပ	•	•	Я	ط	Н	1	1	•	•	٨	•	E	I	A	•	>	•	•	>
(3)	-	^	S	S	•	•	Я	•	Н	L	1	•	•	>	Ξ	ш	•	٠	•	•	•	•	•
(A)61				•		>	•	•	•	•	•	•	•	•	•	•	•	A	•	•	٧	•	•
L8(A)	•				•	>	•	•	•	•	•	•	z		•	•	٠	•	•	•	•	Ь	•
L6(A)		•			•	•	•	•	•	•	•	•			•	•	•	•	•	٠	•	•	•
L2(A)						•		•		•	•		•	•	•	•	•	•	N	•	•	•	•
(A) 11				•		>	•	•	•	•		•	•	•	•	•	•	•	•	•	>	•	٠
Cons-p9	.	ı		α.	<u> </u>	A	H	S	Ö	L.	A	S	S	A	>	エ	>	^	>	-		o	А
L fragment	384	386	388	390	391	392	394	405	434	438	444	450	458	466	474	476	496	524	536	580	622	673	783
	5	7	_1_		•	<u> </u>				·····	•	•											/d

FIG. 7B

16/49

—т		Г	Τ	Т	Т	7		Τ	Τ	Τ	Τ	Т	٦	_	Т	Т		Γ	Τ	Τ	Т	7		Γ	Τ	٦
Cons-F	ပ	>	7	4>	> <	τ .	A,D	α	_		7	בי,		٠		7		4	4	2	د ا		> -			
Cons-D	۰	•		•	۰	0	Q			•	_ ;	H,O	•		, 	a	0	۰	,		0	۰	•	•		•
L4(C)		,		•		•	í	•	•	•	•	۰	•		•	•	۰			•	۰		Н		•	•
L10(B)			,			•	_	,	۰	- -	-	Ŧ	۰		•	٥	•		,	•	٠	S			•	•
(B) _* (7)	v	7		¥	ь	٠	٢	2		•	_	Τ	•		•		٠				٠				•	
L3(B)	Ţ,	•	_	٠	•	۰	٥	3	۰			I	°			٠	٠				٠	S			•	
(A)61			۰	٥	_	۰		•	٥	۰	۰	•	۰		•	۰	,		•	•	•	٠				•
L8(A)		,	٠	٠	I				S	۰	•	•		,	•	٠	,	,			•			•	٠	•
L6(A)		•	0	•		>	•			۰		•		•		_		•		•	•			•		•
12(A)	- 1	•		۰	_		•	۰	٥	•	•	۰		٥	•	٠	ر				۰				٠	а.
(A) 11		۰	•				•	۰	۰	٠			;	×	~				•	^	Z			•	A	۰
Pu-suc)	50 E	C	2	~	>	>	A	V	d.	A	>	. c	,		ပ	S	5 .		A	A	¥	<u>-</u>	- ;	> -	 	
fraamont	ר וומלוווכווו	820	85.7	700	034	+56	937	978	1028	10.31	1043	1067	1001	1097	1188	1915	0171	1223	1226	1339	1300	1503	1303	1528	1535	1662
		VSV	7	٠					NC.3	2																NS4A

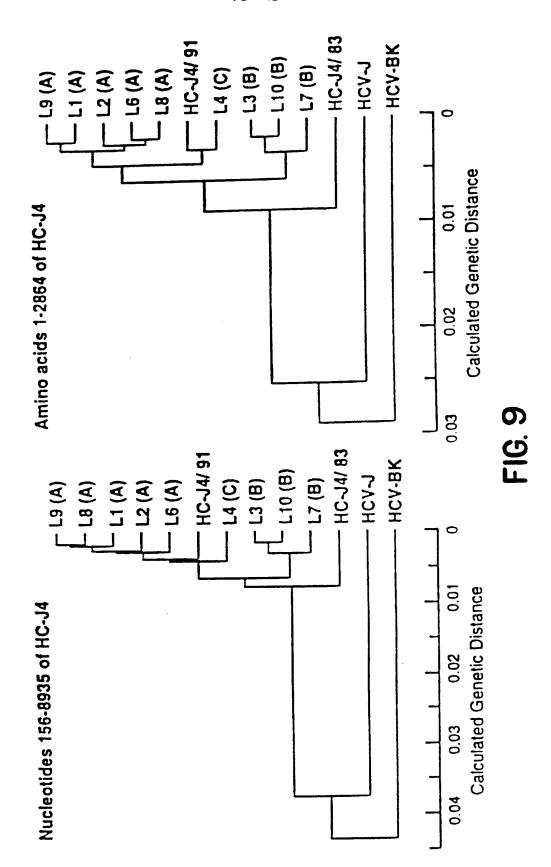
FIG. 70

Cons-F	¥	H,N	S	×	¥	A T	-	_]	J	E,D	۸	L,Q	\	Z	A	ပ		A		S	S	O	S
Cons-D	•	Z		•	•	TA	01	•	•	E,D	•	L,Q	•	•	•	•	•	•		•	٠	•	•
L4(C)		z	۵.		~		,	•	•	•	٠	•	Ξ	•	•	•	•	•	Se	•	9	S	•
L10(B)	•	z				4		•	•	۵	•	0	•	•	•	•	•	٠	S4	•	9		
(8) [7]	•	•	•	•	•	\ \ \	2	•	٠	D	•	0	•	S	•	•	•	•	S10	S	9	•	•
L3(B)	•	z	•	-		<	< 0	٦	4	۵	I	O	•	•	•	•	•	>	8S	S	ဌ	•	•
(A)6J	•	•	•	>	1	1	•	•	•	•	•	•	•		•	•	•	•	LS S	•	•	•	•
L8(A)				1			·	•	•	•	•	•		٠		•	٠	•	S3	•	•	•	F
L6(A)	•	Z	•	•			•	•	•	•	•	•	•		•		•	٠	S2	•	•	•	<u>L</u>
(A)	2	1		1			•	•	•	•	•				•	œ	>	•	S	•	•	•	LL.
L1* (A)							•	•	•		•		•	•	•	•	•	•	\$5			•	•
Cons-p9	×	۵ =	= 0	2	Σ	٠		1		ليا	>		<u></u>	Z	A			A		ی	O.	O	S
L fraament	1753	1805	1000	1949	C017	2136	2146	2226	2259	2262	2334	2371	2385	2692	2757	2785	2824	7861	S fraament	2968	2975	2978	2999
		NS4B L	1		NS5A									NCSD		F							

FIG. 7D

7	11 (4)	12 (4)	16 (A)	18 (A)	(A)	L3 (B)	(8)	L10 (B)	L4 (C)	HC-J4/91HC-J4/83	HC-J4/83
8	(2)			0.75	0.33	- 50	1.53	1.46	0.95	0.83	1.79
(A)	/	0.20	00.0	0.00	50	3					
12 (A)	0.59		0.55	0.35	0.50	1.49	1.51	1.45	0.98	0.82	1.//
3		6,0		0.31	0.55	1.33	1.38	1.29	0.80	0.68	1.58
(A)	0.32	0.42						90	07.0	0.65	1 63
(A)	0.42	0.38	0.31		0.31	1.32	1.54	1.28	0.79	0.00	70.1
(₹) 61	0.35	0.52	0.45	0.35		1.42	1.42	1.38	0.91	0.75	1.66
	1 47	1 43	1.15	1.33	1.36		0.61	0.30	1.43	0.30	1.51
(0)	} -	2					1	27.7	1 17	20.0	154
(17 (8)	1.36	1.33	1.05	1.22	1.22	99.0		0.5/	74.	G.90	5:
110 (B)	1.36	1.33	0.59	1.22	1.26	0.31	0.56		1.37	0.85	1.42
5 2 4		0.80	0.59	0.63	1.26	1.12	1.08	1.01		0.76	1.73
HC - 14 /91		0.91	0.63	0.80	0.87	0.77	0.73	99.0	0.52		1.22
28/ 17 31		1 80	1 68	1.85	1.82	1.75	1.61	1.61	1.71	1.40	
0/+0-01		5	3321								

F1G. 8



SUBSTITUTE SHEET (RULE 26)

		20/49		: :
486 DQRPYC				• • •
YTKPNSSE		ы ы :		ESG.R E.D.P HVR2
468 GWGPIT 1		. «		• • • • • • • • • • • • • • • • • • • •
413 QKIQL 				<u>O</u>
			Δ	F. P. S. P.
ETHTTGRVAGHTTSGFTSLFSSGAS				R
RVAGHTT		> > > > > > > > > > > > > > > > > > >	G R G R G R G R G R G R G R G R G R	GA . 9
ETHTTGRA			T.Y.S.G. T.Y.S.GA. T.Y.S.G T.Y.S.G A.Y.S.G	K.Y.S.
379 AGVDG		: : :		• • •
-	HC-J4L1 (A) : HC-J4L8 (A) : HC-J4L9 (A) : HC-J4/91-21 :	HC-J4L4 (C) : HC-J4/91-23 : HC-J4/91-22 :	HC-J4L7 (B) : HC-J4L10(B) : HC-J4L3 (B) : HC-J4/91-26 : HC-J4/91-25 : HC-J4/91-24 :	HC-J4/91-27 HC-J4/83

90 GGC GACACTCCAC CATAGATCAC TCCCCTGTGA GGAACTACTG TCTTCACGCA GAAAGCGTCT AGCCATGGCG	GGGAGAGCCA TAGTGGTCTG CGGAACCGGT GAGTACACCG GAATTGC	CTCAATGCCT GGAGATTTGG GCGTGCCCCC GCGAGACTGC TAGCCGAGTA GTGTTGG	341 341 5TACT GCCTGATAGG GTGCCCCGGG AGGTCTCGTA GACCGTGCAC C 5. 146 5. 147 5. 148 5. 149 5. 14		poly U-UC region 3' variable region	9513 CCACT CCAGGCCAAT AGGCCTTC CTG poly (U-UC) _n GGTGGCT CCATCTTAG T.:A.NATT. poly (U-UC) ₈₁ AAT A.T.:A.TT. poly (U-UC) ₈₁ AAT Bfr 1	3' conserved region (Cont.)	9514 CCCTAGTCAC GGCTAGCTGT GAAAGGTCCG TGAGCCGCAT GACTGCAGAG AGTGCTGATA CTGGCCTCTC TGCAGATCAT GT	
AC	91 IC-J4 :ITAGTAIGAG IGTCGIGCAG CCTCCAGGAC CCCCCTCCC >CV-J4L6S:	ე: :	271 HC-J4 :GCGAAAGGCC TIGTGGTACT GCCTGATAGG pCV-J4L6S:	3' Untranslated Region	3' variable region	9372 HC-J4 :TGAACGGGA GCTAACCACT CCAGGCC pCV-J4L6S:G.TT.G .GAG.C.T			

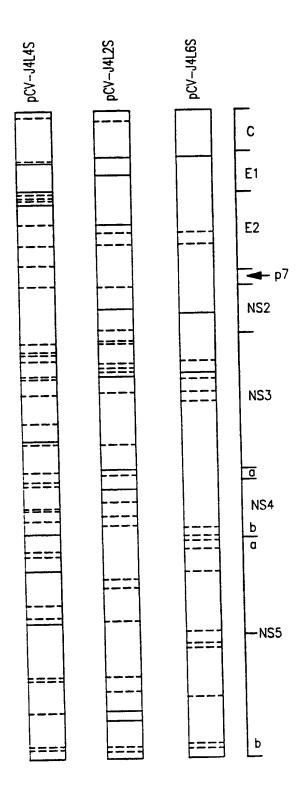


FIG. 12

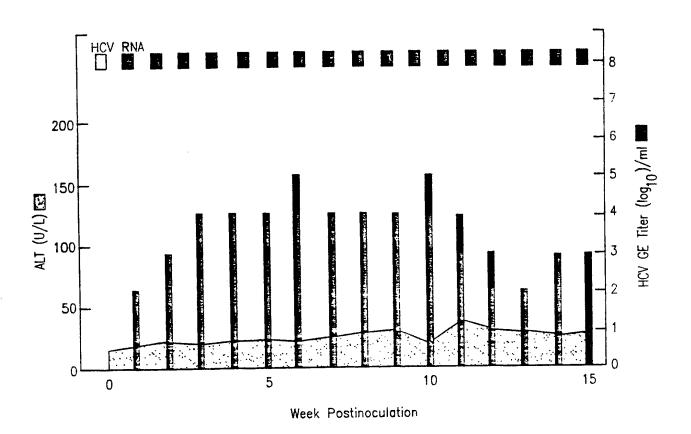


FIG. 13

HC-J4

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GCCAGCCCCC	TGATGGGGGC	GACACTOCAC	CATGAATCAC	TCCCCTGTGA	50
GGAACTACTG	TCTTCACGCA	GAAAGOGICI	AGCCATGGCG	TTAGTATGAG	100
	CCTCCAGGAC				150
	GAGTACACCG				200
	CTCAATGCCT				250
TAGCCGAGIA	GIGITGGGIC	GCGAAAGGCC	TIGIGGIACT	GCCTGATAGG	300
GIGCTIGCGA	GIGCCCCGGG	AGGICICGIA	GACCGIGCAC	CATGAGCACG	350
	CTCAAAGAAA				400
GGACGICAAG	TICCCGGGGG	GIGGICAGAT	CGTTGGTGGA	GITTACCIGI	450
TGCCGCGCAG	GGGCCCCAGG	TIGGGIGIGC	GCGCGACTAG	GAAGGCTTCC	500
GAGCGGTCGC	AACCICGIGG	AAGGCGACAA	CCTATCCCAA	AGGCTCGCCG	550
ACCCGAGGGC	AGGGCCTGGG	CTCAGCCCGG	GIACCCITGG	CCCCICIAIG	600
GCAATGAGGG	CCTGGGGTGG	GCAGGATGGC	TCCTGTCACC	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	650
CGGCCTAGIT	GGGGCCCCAC	GGACCCCCGG	CGTAGGTCGC	GIAACIIGGG	700
TAAGGICATC	GATACCCTTA	CATGCGGCIT	CCCCGATCTC	ATGGGGTACA	750
	accecacac				800
GELGICCGGG	TICIGGAGGA	CGGCGTGAAC	TATGCAACAG	GGAACTTGCC	850
				TGITTGACCA	900
	CGCTTATGAA				950
				CCCACCICAT	1000
				AACAGCICCC	1050
				TGCCAGCGIC	1100
CCCACTACGA	CAATACGACG	CCACGTCGAC	TICCICGIIG	GGACGGCTGC	1150
TTTCTGCTCC	GCTATGTACG	TGGGGGATCT	CIGCGGATCT	ATTITICCTCG	1200
				AGTGCAGGAC	1250
				GCATGGCTTG	1300
				GIGICGCAGT	1350
				GGCCCACTGG	1400
				ACIGGGCIAA	1450
				GAGACCCACA	1500
				GICCCITTIC	1550
				ACGGCAGCTG	1600
				CAAACIGGGT	1650
				CGGGIGCCCG	1700
				AGGGGTGGGG	1750
				CCTTATIGCT	1800
				GCAGGIGIGI	1850
GGTCCAGTGT	ATIGITICAC	CCCAAGCCCI	GIIGIGGIGG	GGACCACCGA	1900

FIG. 14A

HC-J4

10 20	30	40	50	
-		1234567890	1234567890	
TOGITOCOGI GICCCIACGI				1950
TOCTOCTOAA CAACACGOGT	CCCCCACAAG	GCAACIGGIT	CGGCTGTACA	2000
TGGATGAATA GTACTGGGTT	CACTAAGACG	TECEGAGGIC	CCCCGTGIAA	2050
CATCGGGGG GICGGIAACC	GCACCTIGAT	CTGCCCCACG	CACTCCTTCC	2100
GGAAGCACCC CGAGGCTACT	TACACAAAAT	GIGGCIGGG	CCCTCCTTC	2150
ACACCIAGGI GCCIAGIAGA	CTACCCATAC	AGGCTTTGGC	ACTACCCIG	2200
CACICICAAT TTTTCCATCT	TTAAGGTTAG	CATCIATGIG	GGGGGGGIGG	2250
AGCACAGGCT CAATGCCGCA	TGCAATTGGA	CTCGAGGAGA	GCCCIGIAAC	2300
TIGGAGGACA GGGATAGGIC	AGAACTCAGC	COCCTCCTCC	TGICIACAAC	2350
AGAGIGGCAG ATACIGCCCI	GIGCITTCAC	CACCCIACCG	GCTTTATCCA	2400
		TGGACGIGCA		2450
		ATCAAATGGG		2500
		CCICICICCC		2550
		CCTTAGAGAA		2600
		GTATICICI		2650
		CAGGCIGGCI		2700
CGIATCCITT TTATCCCGIA	TGGCCGCIGC	TOCIGOTOCI	ACTGGCGTTA	2750
CCACCACGAG CITACGCCIT	C GGACCGGGAC	ATGGCTGCAT		2800
TECEGITCIT GTAGGICIC			TACTACAAAG	2850
TGTTTCTCAC TAGGCTCATA	A TOGICGITAC	AATACTTTAT		2900
GAGGCGCACA TGCAAGTGT	ediacccc		: GGGGAGGCCG	2950
	A CCICICOCCI		TIAATTITIG	3000
ACATCACCAA ACICCIGCI	C GCCATACICO	GCCCGCTCAI		3050
	C GIACITOGIO		GCTCATTCG	3100
			GICCAAAIGG	3150
TCTTCATGAA GCTGGGCGC	G CIGACAGGI	A CGTACGITIZ	TAACCAICIT	3200
ACCCCACTGC GGGACTGGG	C CCACGCGGG	CIACGAGAC	TIGCGGIGGC	3250
GGTAGAGCCC GTCGTCTTC	T CCGCCATGG	A GACCAAGGI	AICACCIGGG	3300
GAGCAGACAC CGCTGCGTG	T GGGGACATC	A TCTIGGGIC	r Accegierce	3350
GCCCGAAGGG GGAAGGAGA	T ATTTTTGGG	A CCGGCIGATA	A GICICGAAGG	3400 3450
GCAAGGGIGG CGACICCII	G CGCCCATCA	C GGCCLACIU	CAACAAACGC	3 4 50
GGGGGIACT TGGTTGCAT	C ATCACTAGO	C TCACAGGCU	S GLACAALAAL	3550
CAGGICGAAG GGGAGGIIC	A AGIGGITIC	T ACCGCAACA	AAICITICCI	3600
GGCGACCIGC ATCAACGGC	G TGIGCIGGA	C TGICTACCA	T GAGGIGALI.	3650
CGAAGACCCT ACCCCGTCC	A AAAGGICCA	A TUALUUAAA	T. GTALTALCANT.	3700
GTAGACCTGG ACCTCGTCC	G CIGGCAGG		G COCOCICCAI	3750
GACACCATGC AGCTGTGGC	A GCICGGACC	T TIACTIGGI	C WOODSHING	3800
CIGATGICAT TCCGGIGC	SC CGGCGAGGC	LELA ELA DA	JENNITENNA O	2000

FIG. 14B

HC-J4

10 20	30	40	50	
1234567890 1234567890		1234567890	1234567890	
TCCCCCAGGC CCGICICCIA	CCTGAAAGGC	TOCTOGGGIG	GICCATIGCT	3850
TICCCCTICG GGGCACGICG	TEGECETCIT	COCCECTION	GIGIGCACCC	3900
GGGGGTCGC GAAGGCGGTC	GACTICATAC	CCGTTGAGTC	TATGGAAACT	3 95 0
ACCATGOGGT CICOGGICTI	CACAGACAAC	TCAACCCCC	CCCCTCTACC	4000
GCAGACATIC CAAGIGGCAC	ATCTGCACGC	TOCIACIGGO	AGCGGCAAGA	405 0
GCACCAAAGT GCCGGCTGCC	TATGCAGCCC	AAGGGTACAA	GETECTOETC	4100
CTGAACCCGT CCGTTGCCGC	CACCTTAGGG	TITIGGGGGGT	ATATGICCAA	4150
GCACACGGT ATCGACCCT	ACATCAGAAC	TGGGGIAAGG	ACCATTACCA	4200
CGGGCGGCTC CATTACGTAC	TOCACCIATE	GCAAGITCCT	TECCEACCET	4250
GCIGITCIG GGGGGCCIV	TGACATCATA	ATATGIGATG	AGTGCCACTC	4300
AACTGACTOG ACTACCATO	TGGGCATCGG	CACAGICCIG	GACCAAGCGG	43 50
ACACCECTICG ACCCCCCCI	GICGIGCICG	CCACCGCTAC	ACCICCGGGA	44 00
TOTALTITACOG TGCCACACCO	CAATATOGAG	GAAATAGGCC	TGICCAACAA	44 50
TICENCACATO COCTICIATO	G GCAAAGCCAI	CCCCATIGAG	GCCATCAAGG	4500
COCCACION TOTCATITI	TGCCATTCCA	AGAAGAAATG	TCACCAGCIC	4550
CYTTANACT TGACAGGCC	r cogacigaac	GCIGIAGCAI	ATTACCGGGG	460 0
CTITICATICITY TCCGTCATA	C CCCCTATCCC	AGACGICGII	' GICGIGGCAA	4650
CAGACTUT AATGACGG	T TICACCGGCC	ATTTTGACTO	AGTGATCGAC	470 0
TOTALTACAT GTGTCACCC	A GACAGIOGA	TTCAGCTTGG	ATCCCACCI'I'	4750
CACCATTGAG ACGACGACC	G TGCCCCAAG?	A COCCOTOTICE	CGCICGCAAC	4800
COCCACETAG AACTGGCAG	G GGTAGGAGI(GCATCTACAC	GITIGIGACI'	4850
CACCACAAC CCCCTCCC	G CATGITCGAT	r Terreggiee	TGIGIGAGIG	4900
CTATEACTC GGCIGIGCI	T GGTATGAGC	r cacgocogci	GAGACCICGG	4950
TTAGETTECE GECTTACCI	A AATACACCA	GETTECCCET	CIGCCAGGAC	5000
CATCHICEACT TOTGGGAGA	G CGICTICAC	A GGCCTCACCC	ACATAGAIGC	5050
CACTICATE TOCCAGACT	A AACAGGCAG	G AGACAACTT	CCTTACCIGG	5100
TGGCATATCA AGCTACAGI	G TGCGCCAGG	G CICAAGCIC	ACCICCAICG	5150
TOTAL TOTALA	G TCTCATACG	G CTGAAACCT	A CACIGCACGG	5200
GCCAACACCC CIGCIGIAS	TA GGCTAGGAG	C CGTCCAAAA	r GAGGICATICC	5250
TCACACACCC CATAACTA	A TACATCATE	G CATGCATGI	C GGCIGACCIG	5300
GAGGICGICA CTAGCACC	ic ceiecieei	A GCCGGAGIC	C TIGCAGCITT	5350
COCCEPTAC TOCCIGAC	EA CAGGCAGIG	T GGTCATIGIO	G GGCAGGAICA	5400 5450
TOTAL CARCOLAG	T GIOGITOCO	G ACAGGGAAG	r cciciaccag	5450 5500
GAGTICGATG AGATGGAA	GA GIGIGOCIO	A CAACIICCI	T. ACATUGAGCA	5550 5550
GGGAATGCAG CTCGCCGA	SC AATTCAAGC	'A AAAGGCGCT	C GRELIATION	5600
AAACGGCCAC CAAGCAAG	OG GAGGCIGCI	G CICCCGIGG	T. GRAPICCAAP	= :
TGGCGAGCCC TIGAGACC	TT CTGGGCGAA	G CACATGIGG	A ATTICATOR	
CGGAATACAG TACCTAGC	AG GCITATICC	AC TOTACIA	M MALLICUSCUM	2,00

FIG. 14C

HC-J4

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TAGCATCATT	GATGGCATTT	ACAGCITCIA	TCACTAGCCC	GCTCACCACC	5750
CAAAACACCC	TCCTGTTTAA	CATCTTGGGG	GGATGGGTGG	CIGCCCAACT	5800
CCTCCTCCC	AGCGCIGCGI	CAGCTTTCGT	GGGCGCCGGC	ATCGCCGGAG	5850
CGCCTGTTGG	CAGCATAGGC	CTTGGGAAGG	TGCTCGTGGA	CATCTTGGCG	5900
GGCTATGGGG	CAGGGGTAGC	CGGCGCACTC	GIGGCCITIA	AGGICATGAG	595 0
CGGCGAGGIG	CCCTCCACCG	AGGACCIGGT	CAACITACIC	CCTGCCATCC	6000
TCTCTCCTGG	TECCCIEGIC	GICGGGGICG	TGTGCGCAGC	AATACTGCGT	6050
CCCCACCICC	GCCCGGGAGA	GGGGGCTGTG	CAGIGGAIGA	ACCGCCIGAT	6100
ACCGTTCCCT	TOGOGGGGIA	ACCACGICIC	CCCTACGCAC	TATGIGCCIG	6150
AGAGCGACGC	TGCAGCACGI	GICACICAGA	TCCTCTCTAG	CCTTACCATC	6200
ACTCAACTGC	TGAAGOGGCT	CCACCAGIGG	ATTAATGAGG	ACIGCICIAC	625 0
GCCATGCTCC	GCTCGTGGC	TAAGGGATGT	TICCCATICC	ATATGCACGG	6300
TGITGACIGA	CITCAAGACC	TGGCTCCAGT	CCAAACTCCT	GCCGCGGTTA	6350
CCGGGAGTCC	CITICCIGIC	ATGCCAACGC	GGGTACAAGG	CACICICCC	6400
GGGGGACGGC	ATCATGCAAA	CCACCIGCCC	ATGCGGAGCA	CAGATOGCCG	6450
GACATGICAA	AAACGGTTCC	ATGAGGATCG	TACCCCTAC	AACCIGCAGC	6500
AACACGIGG	ACGGAACGIT	CCCCATCAAC	GCATACACCA	CCCCACCTIC	6550
CACACCCTCC	COGGGGGCCA	. ACTATICCAG	GGCGCTATGG	CCCCTC	660 0
CIGAGGAGIA	CGIGGAGGII	ACCCGIGICG	GGGATTTCCA	CIACGIGACG	6650
GGCATGACCA	A CTGACAACGI	' AAAGTGCCCA	TGCCAGGITC	CGGCCCCGGA	670 0
ATTCTTCACC	GAGGIGGATO	GAGTGCGGTT	GCACAGGIAC	GCICCGGGGI.	6750
GCAAACCIC	 	GACGICACGI			6800
TACTIGGIC		, CCCAIGCGAG			6850
GCTTACTTC	ATGCTCACCG				6900
AGCGTAGGC.				CICATCAGCI	6950
AGCCAGTIG.				CCCACCATGA	7000
	C GCTGACCTC				705 0
	A CATCACICGO				7100
				AGATATOCGT	7150
				GCGTTGCCCA	7200
				CIGGAAGGAC	
				CACCTACCAA	7300
				GICCIGACAG	7350
				A GACCTICGGI	7400
				CCCTTCCTGA	
				TOGTACICCT	7500
				r cagogaogg	7550
TCTTGGTCI	A CCGIGAGIC	A GGAGGCTAG	r gaggatgico	G TETGETGETC	7600

FIG. 14D SUBSTITUTE SHEET (RULE 26)

HC-J4

10 20	30	40	50	
10 20 1234567890 1234567890				
AATGICCIAT ACGIGGACAG	1234307630	CACCCATTC	1234307630	7650
AAAGTAAGCT GCCCATCAAC				7700
AACATGGTCT ACGCCACAAC				7750
AACATGGICT ACGCACAAAC GGICACCITT GACAGATTGC				7800
CALACCITT CALACTIC TCAAGCACAT CAAGCCCAAG				7850
TCAAGGAGAG CCTGCAAGCT				7900
ATAGAGGAGG CCIGLAAGGICG				7950 7950
TGGCIAIGG GLAAAGGALG	TOGGARCI	ATCOMO	AACACCAATT	8000
ACATCCCCTC CGTGTGGGAG		WHENT TON	ANCORONA	8050
CACACCACCA TCATGGCAAA	TIEDERA	TICIGOSICC	WATCHIENE.	8100
GGCAGGCCCC AAGCCAGCTC	GCTIAICGI	ATTCCCALAC	CIGGGGGIIC	8150
GIGIAIGCGA GAAGAIGGCC				8200
GCCGIGATGG GCTCCTCATA	CGGATTICAA	TACTUCCUA	AGLAGLGGGT	8250
CGAGITCCIG GIGAATACCI	GGAAAICAAA	GAAATGCCCT	AIGGGITTCT	8300
CATATGACAC CCGCIGITTI				
GITGAGGAGT CAATTTACCA				8350
GGCCATAAGG TCGCTCACAG	AGCGGCTTTA	CATCGGGGG1	CCCCIGACIA	8400
ACTCAAAAGG GCAGAACTGC	GGITATCGCC	GGIGCCGCGC	AAGIGGGIG	8450
CTGACGACTA GCTGCGGTAA	. TACCCICACA	TGITACITGA	AGGCCACIGC	8500
AGCCIGICGA GCIGCAAAGC	TOCAGGACTG	CACGATGCTC	GIGAACGGAG	8550
ACGACCITGT CGITATCIGI	, GYYYGCGCGG	GAACCCAGGA	GGATGCGGCG	8600
CCCTACGAG CCTTCACGGA	. GGCIAIGACI	AGGIATICCG	cccccccc	8650
GGATCOGCCC CAACCAGAAT	ACGACCIGGA	. GCTGATAACA	TCATGITICCI	8700
CCANTGIGIC AGICGCGCAC	GATGCATCIC	GCAAAAGGGI	ATACIACCIC	8750
	CCTTGCACGG			8800
ACACACTOCA ATCAACTOTT	GGCTAGGCAA	A TATCATCATO	TATGCGCCCA	8850 -
CCCIATGGGC AAGGATGATI	CIGATGACIO	ACTITITICIO	CATCCITCIA	8900
GCTCAAGAGC AACTTGAAAA	A AGCCCIGGAI	TGICAGAICI	ACGGGGCTTG	8950
CTACTOCATT GAGOCACTIO	ACCIACCICA	A GATCATIGAA	CGACTCCATG	9000
GICTIAGCGC ATTTACACTO	CACAGITACI	CICCAGGIGA	CATCAATAGG	9050
GIGGCTICAT GCCTCAGGA	A ACTIGGGGIA	A OCACOCTIGO	CAACCIGGAG	9100
ACATOGGGOC AGAAGIGIO	GCGCTAAGC	r ACIGICCCAC	GGGGGGAGGG	9150
COGCCACTIG TGGCAGATA	CICITIAAC.	r gggcagiaac	GACCAAGCIT	9200
AAACTCACTC CAATCCCGG	CGCGTCCCA	G CIGGACTIG	CIGGCIGGIT	9250
CGTCGCTGGT TACAGCGGG	G GAGACATAT	A TCACAGOCTO	TCICGIGCCC	9300
GACCCCCCIG GITTCCGIT	G TOCCIACIO	TACITICIG	r aggggiaggc	9350
ATTTACCICC TCCCCAACC	G ATGAACGGG	G AGCTAACCA	TOCAGGCCTT	9400
AAGCCATTIC CIGITITIT	T TTTTTTTT	r trrtrrrrr	r TCTTTTTTT	9450
TREFFICER TREEFICIT	I TITTOCTIA	C TITTICCCI	r CTTTAATGGT	9500

FIG. 14E

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GGCTCCATCT	TAGCCCTAGT	CACGGCTAGC	TGTGAAAGGT	CCGIGAGCCG	9550
CATGACIGCA	CACACICCIC	ATACTGGCCT	CICIGCAGAT	CAIGI	9595

FIG. 14F

10 20 30 40 50	
1234567890 1234567890 1234567890 1234567890 1234567890	<u> </u>
MSINPKPORK TKRNINRRPO DVKFPGGGQI VGGVYLLPRR GPRLGVRATR	50 100
KASERSOPRG RROPIPKARR PEGRAWAOPG YFWPLYGNEG LGWAGWLLSP	100
RGSRPSWGPT DPRRRSRNLG KVIDILITOGF ADIMGYIPIV GAPLGGAARA	150
LAHGVRVLED GVNYATGNLP GCSFSIFLLA LLSCLTIPAS AYEVRWSGI	200
YHVINDOSNS SIVYFAADVI MHIPGOVPOV QEGNSSROW ALITPILAARN	250
ASVPITTIRR HVDLLWGIAA FCSAMYVGDL CGSIFLVSQL FIFSPRRHET	300
VODONOSIYP CHVSCHRMAW DMMNWSPIT ALVVSQLLRI PQAVVDMVAG	350
AHWGVLAGLA YYSMVGWAK VLIVALLFAG VDGEIHITGR VAGHTTSGFT	400
SLFSSCASOK IOLVNINGSW HINRIALNON DSLOTGFFAA LFYAHKFNSS	45 0
GCPERMASCR PIDWFAQGWG PITYTKPNSS DQRPYCWHYA PRPCGVVPAS	500
QVCGPVYCFT PSPVVVGTTD RSGVPTYSWG ENETDVMLLN NTRPPQGWF	550
GCIMMISIGF TRICGGPPCN IGGVENRILI CPIDCFRKHP EATYIKCGSG	600
PWLTPRCLVD YPYRLWHYPC TLNFSIFKVR MYVGGVEHRL NAACNWIRGE	650
RONLEDRORS ELSPILLSTT EWQILPCAFT TLPALSIGLI HLHQNIVDWQ	700
YLYGVGSAFV SFAIKWEYIL LLFLLLADAR VCACLWMMLL IAQAFAALEN	750
INVINAASVA GAHGILSFLV FFCAAWYIKG RLAPGAAYAF YGWPLLLLL	800
LALPPRAYAL DREMAASOGG AVLVGLVFLT LSPYYKVFLT RLIWWLQYFI	850
TRAFAHMOW VPPINVRGGR DAIILLTCAV HPELIFDITK LLLATIGPLM	900
VIOAGTTRVP YEVRAOGLIR ACMLVRKVAG CHYVQMVEMK LGALIGIYVY	950
NHITPIROWA HAGIRDIAVA VEPVVFSAME TKVITWGADI AACGDIIIGL	1000
PASARROKET FLOPADSLEG OGWRLLAPIT AYSOQIRGVL GCIITSLIGR	1050
DENOVECENO VVSTATOSFL ATCINGVOWT VYHGAGSKTL AGPKGPITOM	1100
YTM/DIDING WOAPPGARSM TPCSCGSSDL YLVTRHADVI PVRRRGDSRG	1150
STISPREVSY IKGSSGGPLL CPSCHVVGVF RAAVCTRGVA KAVDFTPVES	1200
METTMRSPVF TDNSTPPAVP QIFQVAHLHA PIGSGKSTKV PAAYAAQGYK	1250
MAMINESVAA TIGEGAYMSK AHGIDENIRT GVRTITIGGS ITYSTYCKEL	1300
ADOCCOCAY DITTOPPORTS TOSTTILGIG TVLDQAETAG ARLVVLATAT	1350
PROSVIVEHE NIFEIGLSNN GEIPFYGKAI PIFAIKGGRH LIFCHSKKKC	1400
DELAKUTEL GINAVAYRG LDVSVIPPIG DVVVVATDAL MIGFIGDEDS	1450
VIDOMOVIO TVDFSLDPIF TIEITIVPQD AVSRSQRRGR TGRGRSGIYR	1500
FYTEGEREGG MEDSSVICEC YDAGCAWYEL TPAETSVRLR AYLNINGLEV	TOOU
CODHIFFMES VETGLITHIDA HFLSOIKQAG DVFPYLVAYQ ATVCARAQAP	TP00
PPSWDOMNKC LIRLKPILLIG PIPLLYRLGA VQNEVILITHP TIKYIMACMS	1650
ADIFATETW VINCEVIALL AAYCLTIGSV VIVERIILSG KPAVVPLREV	7,00
LYOFEDEMEE CASOLPYIED CMOLABOFKO KALGLIQIAT KOAFAAAPVV	1/50
ESKAPATETE MAKHMANETS GIOYLAGIST LPGNPATASI MAETASITSP	TROO
LITONTLEN ILGGWAAOL APPSAASAFV GAGIAGAAVG SIGLGKVLVD	7820
ILAGYGAGVA GALVAFKVMS GEVPSTEDLV NILPAILSPG ALVVGVVCAA	1900

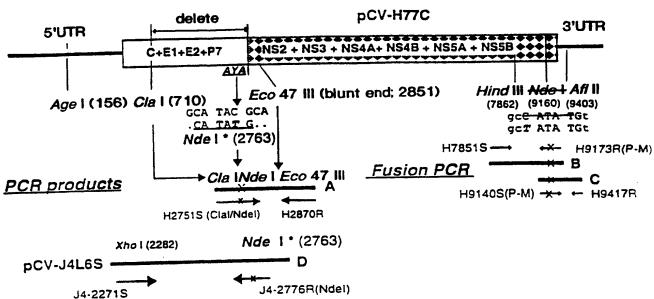
FIG. 14G

SUBSTITUTE SHEET (RULE 26)

10	20	30	40	50	
1234567890 12345	67890 12	34567890	1234567890	1234567890	
TI RRHVGPGE GAVOV	MINRLI AF	ASRGNHVS	PIHYVPESDA	AARVIQILSS	1950
LITTOLLKRL HOWIN	VEDCST PC	SGSWLRDV	WDWICIVLID	FKIWLQSKLL	2000
PRLPGVPFLS CORG	KGVWR GD	GIMOTTCP	CCAQIACHVK	NGSMRIVGPR	2050
TOSNIWHGIF PINA	YTTGPC TP	SPAPNYSR	ALWRVAAEEY	VEVIRVGDFH	2100
VATISMITITANV KCPC	WPAPE FF	TEVDGVRL	HRYAPACKPL	LREDVIFQVG	2150
INDYINGSOL PCEP	EPDVIV LI	SMLIDPSH	TTAETAKRRL	ARGSPPSLAS	2200
SSASOLSAPS LKATO	CTIHHD SP	DADLIEAN	LLWRQEMGGN	TIRVESENKV	2250
VII DSFEPIH AEGD	EREISV AA	EILRKSRK	FPSALPIWAR	PDYNPPLLES	2300
WKDPDYVPPV VHGC	PLPPIK AP	PIPPPRRK	RIVVLIESW	SSALAELATK	2350
TFGSSGSSAV DSGI	ATALPO LA	SDDGDKGS	DVESYSSMPP	LEGEPGOPDL	2400
SDGSWSIVSE EASE	DVVCCS MS	YTWICALI	TPCAAEESKL	PINPLSNSLL	2450
RHHNMVYATT SRSA	SLROKK VI	FDRLQVLD	DHYRDVLKEM	KAKASTVKAK	2500
II.STEFACKI, TPPH	SAKSKF GY	CAKDVRNL	SSRAVNHIRS	WEDLLEDIE	2550
TOTTTTTMAK SEVE	CVOPEK GO	RKPARLIV	FPDLGVRVCE	KMALYDVVST	2 60 0
T.POAVMGSSY GFOY	SPKORV EF	IJVNIWKSK	KCPMGFSYDI	RCFDSIVIES	2650
DIRVEESIYO CCDL	APEARO AJ	RSLIERLY	IGGPLINSKG	QVCGYRRCRA	270 0
SGVLTTSCGN TLTC	YLKATA AC	RAAKLQDC	IMIVNGDDLV	VICESAGIQE	275 0
DAAAIRAFTE AMIR	YSAPPG DI	PPOPEYDLE	LITSCSSNVS	VAHDASGKRV	2800
YYTTROPTTP LARA	AWETAR H	PINSWLGN	I IIMYAPILWA	RMILMIHFFS	2850
TTI AOFOLEK ALDO	OIYGAC Y	SIEPLDLPQ	ITERLHGLSA	FILHSYSPGE	290 0
TATEL TALL SCALER I CALL	PERIWR H	RARSVRAKI	, <u>LSQGGRAAT</u> C	GRYLFNWAVR	295 0
TKLKLTPIPA ASQI	DLSGNF V	AGYSGGDIY	HSLSRARPRA	FPLCILLLSV	300 0
GVGIYLLPNR					301 0

FIG. 14H

#2. Strategy for constructing chimeric clone of HCV (pH77CV-J4) which contains the nonstructural region of strain H77 and the structural region of strain HC-J4



- 1. Fragment A, B, C and D; PCR amplification from pCV-H77C or pCV-J4L6S
 - Fragment A; additional Cla I site, artificial Nde I site induced by a single mutation (C→T at nt 2765 of H77C) and authentic Eco47 III site
 - Fragment B and C; eliminated Nde I site by a single mutation within the primers (C \rightarrow T at nt 9158 of H77C) , and fusion PCR with both fragments
 - Fragment D; artificial Nde I site induced by 2 point mutations within the primer (T→A at nt 2762 and C→T at nt 2765 of J4L6S)
- 2. TA cloning of PCR products
- 3. Sequence analysis
- 4. Cloning of Fragment A (Cla I-Eco 47III) and Fragment B/C (Hind III-Afl II) with correct sequence into pCV-H77C
- 5. Complete sequence analysis of new cassette vector [pH77CV], into which the structural regions of different genotypes can be inserted.
- 6. Cloning of Fragment-Age I/Xho I (cut out from pCV-J4L6S) and Fragment D (Xho I-Nde i) with correct sequence into the new cassette vector; 3 piece ligation
- 7. Complete sequence analysis of 1a+1b chimera [pH77CV-J4]
- 8. In vitro transcription (within 24 hours of inoculation)
- 9. Percutaneous intrahepatic transfection into chimpanzee

FIG. 15

GCCAGCCCC TGATGGGGCC GACACTCCAC CATGAATCAC TCCCCTGTGA	50
GGAACIACIG TCITCACGCA GAAAGCGICT AGCCATGGCG TIAGIATGAG	100
TGTCGTCCAG CCTCCAGGAC CCCCCCTCCC GGGAGAGCCA TAGTCGTCTG	150
CCGAACCCCT CACIACACCG GAATTCCCAG GACCACCCCG TCCTTTCTTG	200
CATCAACCCG CICAAIGCCI GGAGATIIGG GCGIGCCCCC GCGAGACIGC	250
TAGCOCAGIA GIGITGGGIC GOCAAAGGOC TIGIGGIACT GCCIGATAGG	300
GIGCIIGOGA GIGCCCCGGG AGGICIOGIA GACOGIGCAC CATGAGCACG	350
AATCCIAAAC CTCAAAGAAA AACCAAACGI AACACCAACC GCCGCCCACA	400
GCACGICAAG TICCCGGGCG GIGGICAGAT CGTIGGIGGA GITIACCIGI	450
TECCECECAG GEGCCCCAGG TIGGGIGIGC GCGCGACIAG GAAGGCTICC	500
CACCCCICC AACCICGICG AAGCCCACAA CCIATCCCAA AGCCICGCCG	550
ACCOGAGESC AGGEOCIGES CICAGOCOGG GIACOCTIGG COCCICIAIG	600
GCAATGAGGG CCTGGGGTGG GCAGGATGGC TCCTGTCACC CCGGGGCTCC	650
CGCCCTAGTT GGGGCCCCAC GGACCCCCGG CGTAGGTCGC GTAACTTGGG	700
TAAGGICATC GATACCCTTA CATGCGGCTT CGCCGATCTC ATGGGGTACA	750
TICCGCICGI CGGCGCCCC CIAGGGGGG CIGCCAGGGC CIIGGCACAC	800
GGIGICCGGG TICIGGAGGA CGGCGIGAAC TAIGCAACAG GGAACITGCC	850
COGTIGCICT TICICIATOT TOCTOTIGGO TOTGCIGICO TGITIGACCA	900
TCCCAGCITC CGCITATGAA GTGCGCAACG TGTCCGGGAT ATACCATGTC	950
ACCAACCACT CCTCCAACTC AACCATTGTG TATGACCCAG CCGACGTCAT	1000
CATGCATACT CCCGGGIGCG TGCCCIGIGT TCAGGAGGGT AACAGCTCCC	1050
GITCCICCO ACCCCICACI CCCACCCICG CCCCACGAA TCCCACCTIC	1100
CCCACTACGA CAATACGACG CCACGTCGAC TIGCTCGTTG GGACGCTGC	1150
TITCICCICC CCIAIGIACG TGGGGGATCT CIGCGGATCT ATTITCCICG	1200
TCTCCCAGCT GITCACCTTC TCGCCTCGCC GGCATGAGAC AGTGCAGGAC	1250
TGCAACIGCT CAATCIATCC CGGCCATGIA TCAGGICACC GCATGGCITG	1300
GGATATGATG ATGAACTGGT CACCTACAAC AGCCCTAGTG GTGTCGCAGT	1350
TECTICOGGAT COCACAAGCT GTOGTGGACA TEGTGGCGGG GGCCCACTGG	1400
GGAGICCIGG CGGCCFIGC CIACIATICC AIGGIAGGA ACIGGCCIAA	1450
GGITCIGATT GIGGCGCIAC TCTTTGCCGG CGTTGACGGG GAGACCCACA	1500
CGACCGGGAG GGIGGCCGGC CACACCACCT CCGGGITCAC GICCCITTIC	1550
TCATCIGGG CGICICAGAA AATCCAGCIT GIGAATACCA ACGCAGCIG	
GCACATCAAC AGGACTGCCC TAAATTGCAA TGACTCCCTC CAAACTGGGT	1650
TCTTTGCCGC GCTGTTTTAC GCACACAAGT TCAACTCGTC CGGGTGCCCG	1700
CACCCATICG CCACCICCCG CCCCATICAC TGGITCGCCC AGGGIGGGG	1750
CCCCATCACC TATACTAAGC CTAACAGCTC GGATCAGAGG CCTTATTGCT	1800

FIG. 16A

GCCATTACGC GCCTCGACCG TGTGGTGTCG TACCCGCGTC GCAGGTGTGT	1850
GGICCAGIGT ATTGITTCAC CCCAAGCCCT GITGIGGIGG GGACCACCGA	1900
TOGITOCGGT GICCCIACGT ATACCIGGG GCAGAATGAG ACAGACGIGA	1950
TGCTCCTCAA CAACACGCGT CCGCCACAAG GCAACTGGTT CGGCTGTACA	2000
TOCATCAATA GIACIGGIT CACTAAGACG TOCGGAGGIC CCCCGIGIAA	2050
CATCGGGGG GICGGTAACC GCACCITICAT CIGCCCCACG GACIGCITICC	2100
GGAAGCACCC CGAGGCTACT TACACAAAAT GIGGCTCGGG GCCCIGGITG	2150
ACACCIAGGI GCCIAGIAGA CIACCCATAC AGGCTTIGGC ACIACCCCIG	2200
CACTOTOAAT TITTCCATOT THAAGGITAG GATGIATGIG GGGGGGGGG	2250
AGCACAGGCT CAATGCCGCA TGCAATTGGA CTCGAGGAGA GCGCTGTAAC	2300
TIGGAGGACA GGGATAGGIC AGAACICAGC CCGCIGCIGC TGICIACAAC	2350
AGAGIGGCAG ATACIGCCCT GIGCTTICAC CACCCTACCG GCTTIATCCA	2400
CIGGITICAT CCATCICCAT CAGAACATCG TGGACGIGCA ATACCIGIAC	2450
GGIGIAGGGT CAGCGTTTGT CICCTTTGCA ATCAAATGGG AGIACATCCT	2500
GITGETTITE CITETECTES CAGACGOGG CGIGIGIGCC TECTIGIGGA	2550
TGATGCIGCT GATACCCCAG GCTGAGGCCG CCTTAGAGAA CTTGGTGGTC	2600
CTCAATGOGG CGTCCGTGGC CGGAGCGCAT GGTATTCTCT CCTTTCTTGT	2650
GITCITCIEC GCCCCCIGGI ACATTAAGGG CAGGCIGGCI CCIGGGGCGG	2700
CGTATGCTTT TTATGGCGTA TGGCCGCTGC TCCTGCTCCT ACTGGCGTTA	2750
CCACCACGAG CATATGCACT GGACACGGAG GTGGCCGCGT CGTGTGGCGG	2800
CGPIGFICIT GICGGGFIAA TGGCGCIGAC TCIGICGCCA TATTACAAGC	2850
CCTATATCAG CTGGTGCATG TGGTGGCTTC AGPATTTTCT GACCAGAGIA	2900
GAAGCGCAAC TGCACGTGTG GGTTCCCCCC CTCAACGTCC GGGGGGGGCG	2950
CGATGCCGIC ATCTIACICA TGIGIGIAGI ACACCCGACC CIGGIATITG	3000
ACATCACCAA ACTACTCCTG GCCATCTTCG GACCCCTTTG GATTCTTCAA	3050
CCAGTTICC TTAAAGICCC CTACTTCGIG CCCGTTCAAG CCCTTCTCCG	3100
CATCIGCGCG CIACCGCGA AGATAGCCGG AGGICATTAC GIGCAAAIGG	3150
CCATCATCAA GITAGGGGG CITACIGGCA CCIATGIGIA TAACCATCIC	3200
ACCCCICTIC GAGACIGGC GCACAACGGC CIGCGAGAIC IGGCCGIGGC	3250
TGTGGAACCA GTCGTCTTCT CCCGAATGGA GACCAAGCTC ATCACGTGGG	3300
GGGCAGATAC CGCCGCGIGC GGIGACATCA TCAACGGCIT GCCCGICICT	3350
GCCCGIAGGG GCCAGGAGAT ACIGCTIGGG CCAGCCGACG GAAIGGICIC	3400
CAAGGGGGG AGGITGCTGG CGCCCATCAC GGCGTACGCC CAGCAGACGA	3450
GAGGCCTCCT AGGGIGIATA ATCACCAGCC TGACTGGCCG GGACAAAAAC	3500
CAAGTGGAGG GTGAGGTCCA GATCGTGTCA ACTGCTACCC AAACCTTCCT	3550
GCCAACGIGC ATCAATGGGG TATGCIGGAC TGICTACCAC GGGGCCGGAA	3600

FIG. 16B

CGAGGACCAT CGCATCACCC AAGGGTCCTG TCATCCAGAT GTATACCAAT	3650
GIGGACCAAG ACCITGIGGG CIGGCCCGCT CCICAAGGIT CCCGCICATT	3700
CACACCCIGI ACCIGCGCCI CCICGCACCT TIACCIGGIC ACGAGGCACG	3750
CCGATGICAT TCCCGTGCGC CGCCGAGGTG ATAGCAGGGG TAGCCTGCTT	3800
TOGOCCOGGC CCATTICCIA CITGAAAGGC TOCICGGGGG GICCGCIGIT	3850
GIGCCCCCC GGACACCCC TGGCCCIATT CAGGCCCCC GIGIGCACCC	3900
GIGGAGIGGC TANAGCGGIG GACTITATCC CIGIGGAGAA CCTAGGGACA	3950
ACCATGAGAT CCCCGGIGIT CACGGACAAC TCCTCTCCAC CAGCAGIGCC	4000
CCAGAGCITC CAGGIGGCCC ACCIGCATGC TCCCACCGGC AGCGGTAAGA	4050
GCACCAAGGT CCCGGCTGCG TACGCAGCCC AGGGCTACAA GGTGFTGGTG	4100
CTCAACCCCT CIGTIGCIGC AACGCIGGGC TTIGGIGCIT ACAIGICCAA	4150
GCCCATGGG GTTGATCCTA ATATCAGGAC CGGGGTGAGA ACAATTACCA	4200
CTRECAGOOC CATCACGIAC TOCACCIACG GCAAGITCCT TGCCGACGCC	4250
CETTECTICAG CAGGICCITA TCACATAATA ATTIGICACG AGICCCACIC	4300
CACCEATICC ACATCCATCT TOGGCATCGG CACTGTCCTT CACCAAGCAG	4350
AGACTECTES GEOGRACIG GITGIECTOG CCACTECTAC CCCTCCGGC	4400
TO CHEACHE TERCECATEE TAACATEGAG GAGGIIGCIC TERCEACEAE	4450
COCAGAGATO COCTITUACE GCAAGGCUAT COCCCIOGAE GIGATCAAGE	4500
CONTAGACA TOTCATOTTO TOCCACTOAA AGAAGAAGTG CGACGAGCTC	4550
COCCEARCE TEGICGEATT GESCATCAAT GEOGIGSEET ACTACESES	4600
TCITGACGIG TCIGICATCC CGACCAGCGG CGAIGITGIC GICGIGICGA	4650
CCGATGCTCT CATGACTGGC TTTACCGGGG ACTTCGACTC TGTGATAGAC	4700
TOCARCAGI GIGICACTCA GACAGIOGAT TICAGOCITG ACCOTACCIT	4750
TACCATIGAG ACAACCACC TCCCCCAGGA TGCTGTCTCC AGGACTCAAC	4800
GCCGGGGCAG GACTGGCAGG GGGAAGCCAG GCATCTATAG ATTTGTGGCA	4850
CCGCGCGACC CCCCTCCGG CATGITCGAC TCGTCCGTCC TCTGTCAGTG	4900
CIATGAGGG GGCIGIGCIT GGIATGAGCT CAGGGGGGC GAGACIACAG	4950
TIAGGCIACG AGCGIACATG AACACCCCGG GGCTTCCCGT GTGCCAGGAC	5000
CARCINGA AN INVICACACIC CENCITUAGE CECCICACIC ATATACATEC	505 0
CACHUTURA TO CACACAA AGCAGAGIGG GGAGAACITT CCITACCIGG	5100
TAGOGIACCA AGCCACOGIG TGCGCIAGGG CICAAGCCCC TCCCCCATCG	5150
TOGGACCAGA TGIGGAAGIG TITGATCOCC CTTAAACCCA CCCICCATGG	5200
GCCAACACCC CIGCIATACA GACIGGGGC TGITCAGAAT GAAGICACCC	5250
TGACGCACCC AATCACCAAA TACATCATGA CATGCATGTC GGCCGACCTG	5300
GAGGICGICA CGAGCACCIG GGIGCICGIT GCCGCGGICC TGCCIGCICT	5350
GECCECGIAT TECCIGICAA CAGECIGCET GETCATAGIG GECAGGATCE	5400
GCCGCIAI IGCCIGIGI. CITTOTTO	

FIG. 16C

The second summaring and second secon	5450
TCTTGTCCGG GAAGCCGGCA ATTATACCTG ACAGGGAGGT TCTCTACCAG	5500
CAGITOGAIG AGAIGGAAGA GIGCICICAG CACITACOGT ACAICGAGCA	5550
AGGGATGATG CICGCIGAGC AGITCAAGCA GAAGGCCCTC GGCCTCCTGC	5600
AGACCGCGTC CCGCCATGCA GAGGITATCA CCCCTGCTGT CCAGACCAAC	
TGGCAGAAAC TCGAGGICIT TIGGGCGAAG CACATGIGGA ATTICATCAG	5650
TGGCATACAA TACTTGGCGG GCCIGICAAC GCTGCCTGGT AACCCCGCCA	5700
TIGCTICATT CATGCCTTTT ACAGCIGCCG TCACCAGCCC ACTAACCACT	5750
GGCCAAACCC TCCTCTTCAA CATATIGGGG GGGIGGGIGG CIGCCCAGCT	5800
CCCCCCCC GGIGCCCTA CIGCCITIGI GGGIGCIGGC CIACCIGGC	5850
CCGCCATCGG CAGCGITGGA CIGGGGAAGG TCCTCGIGGA CATTCITGCA	5900
GGGIATGGGG CGGGCGTGGC GGGAGCTCTT GTAGCATTCA AGATCATGAG	5950
CGGIGAGGIC CCCTCCACGG AGGACCIGGT CAATCTGCTG CCCGCCATCC	6000
TCICGCCIGG ACCCTIGIA GICGGIGIGG TCIGCGCAGC AATACIGGC	6050
COCCACGITG CCCCGCCCA GCCCCAGIG CAATGGATGA ACCGCTAAT	6100
ACCUTICGOC TOCOGGGGA ACCATGITTO COCCACGCAC TACGIGCOGG	6150
AGAGCGATGC AGCCGCCCC GTCACTGCCA TACTCAGCAG CCTCACTGTA	6200
ACCCAGCICC TGAGGGGACT GCATCAGIGG ATAAGCICGG AGIGIACCAC	6250
TOTATION OF GENERAL TAAGGGACAT CIGGGACIGG ATATGCGAGG	6300
TOTTCACTCA CTITAAGACC TGGCTGAAAG CCAAGCTCAT GCCACAACTG	6350
CTICAGATIC CCITTGIGIC CTICCCAGCGC GGGTATAGGG GGGTCTGGCG	6400
AGGAGAGGC ATTATGCACA CICGCIGCCA CIGIGGAGCT GAGAICACIG	6450
CACATGTCAA AAACGGGACG ATGAGGATCG TCGGTCCTAG GACCTGCAGG	6500
AACATGIGGA GIGGGACGIT CCCCATTAAC GCCTACACCA CGGGCCCCIG	6550
TACTOCCCTT CCTGCGCCCA ACTATAAGTT CGCGCTGTGG AGGGTGTCTG	6600
CAGAGGAATA CGIGGAGATA AGGCGGGIGG GGGACTICCA CIACGIAICG	6650
CCUATGACIA CIGACAATCI TAAAIGCCCG TGCCAGATCC CATCGCCCGA	6700
ATTITUTCACA GAATTIGGACG GGGTGCGCCT ACACAGGTTT GCGCCCCCTT	6750
CCAACCCCTT CCTCCCCCAC GACGTATCAT TCACAGTAGG ACTCCACCAG	6800
TACCOCATOR GETCECAATT ACCTIGCGAG CCCGAACCGG ACGTAGCCGT	6850
CHICACETIC ATGCTCACIG ATCCCTCCCA TATAACACA GAGGCGGCCG	6900
CCACAACTIT CCCCACACGC TCACCCCCTT CIATGCCCAG CICCICGCI	6950
ACCIACTET COCCICCATO TOTCAAGGCA ACTIGCACCG CCAACCAIGA	7000
CICCUTTEAC GCCGAGCICA TAGAGGCIAA CCICCIGIGG AGGCAGGACA	7050
TECHNOLOGY CATCACCAGG GITGAGTCAG AGAACAAAGT GGIGATICIG	7100
CACTUCTUCE ATCCCCTTGT CGCAGAGGAG GATGAGCGGG AGGICICCGT	7150
ACCIGCAGAA ATTCIGCGGA AGICTCGGAG ATTCGCCCGG GCCCTGCCCG	7200

FIG. 16D

TCTGGGGGG GCCGGACTAC AACCCCCCGC TAGTAGAGAC GTGGAAAAAG	7250
CCTGACTACG AACCACCTGT GGTCCATGGC TGCCCGCTAC CACCTCCACG	7300
GICCCCICCT GIGCCICCGC CICGGAAAAA GCGIACGGIG GICCICACCG	7350
AATCAACCCT ATCIACTGCC TIGGCCGAGC TIGCCACCAA AAGTTTIGGC	7400
AGCICCICAA CITCCGGCAT TACGGGCGAC AATACGACAA CATCCICIGA	7450
GCCGGCCCT TETGGCTGCC CCCCGACTC CGACGITGAG TCCIATTCIT	7500
CCATGOCCCC CCTGGAGGG GAGCCTGGGG ATCCGGATCT CAGCGAGGG	7550
TCATEGICEA CEGICAGIAG TEGGECCCAC ACEGAAGATE TCGIGIECIE	7600
CICAATGICT TATTCCIGGA CAGGCGCACT CGICACCCCG TGCGCIGGGG	7650
AAGAACAAAA ACTGCCCATC AACGCACTGA GCAACTCGTT GCTACGCCAT	7700
CACAATCIGG IGIATICCAC CACTICACGC AGIGCTIGCC AAAGGCAGAA	7750
GAAAGICACA TITGACAGAC TGCAAGITCT GGACAGCCAT TACCAGGACG	7800
TGCTCAAGGA GGTCAAAGCA GCGCCGTCAA AAGTGAAGGC TAACTTGCTA	7850
TCCGIAGAGG AAGCITGCAG CCTGACGCCC CCACATTCAG CCAAATCCAA	7900
GITTGGCTAT GGGGCAAAAG ACGTCCGTTG CCATGCCAGA AAGGCCGTAG	7950
CCCACATCAA CICCGIGIGG AAAGACCTIC TOGAAGACAG TGIAACACCA	8000
ATAGACACTA CCATCATGGC CAAGAACGAG GITTTCTGCG TTCAGCCTGA	8050
CAACCCCCC CACCTCCCC CACCTCCCCC	8100
TECECGIGIE CEACAACATE COCCIGIACE ACGIEGITAE CAACCICCCC	8150
CIGCOCGIGA TEGGAACCIC CIACGEATIC CAATACICAC CAGGACACCG	8200
GETTGAATTC CTCGTGCAAG CGTGGAAGTC CAAGAAGACC CCGATGGGT	8250
TCTCGIATGA TACCCGCTGT TTTGACTCCA CAGTCACTGA GAGCGACATC	8300
CGTACGGACG AGGCAATTTA CCAATGTTGT GACCTGGACC CCCAAGCCCG	8350
CGIGGCCATC AAGTCCCTCA CIGAGAGGCT TIAIGITGGG GGCCCTCTTA	8400
CCAATTCAAG GGGGAAAAC TGCGGCTACC GCAGGTGCCG CGCGAGCGC	8450
GTACIGACAA CTAGCIGIGG TAACACCCIC ACTIGCIACA TCAAGGCCCG	8500
GECAGCCIGT CGAGCCGCAG GGCTCCAGGA CTGCACCATG CTCGTGTGTG	8550
COCACCACTT AGICGITATC TGICAAAGIG CGGGGGICCA GGAGGACGCG	8600
GCCAGCCIGA CAGCCTICAC GCAGGCIAIG ACCAGGIACT CCGCCCCCCC	8650
CGGGGACCCC CCACAACCAG AATACGACTT GGAGCTTATA ACATCATGCT	8700
CCICCAACGI GICAGICGCC CACGACGCCG CIGGAAAGAG GGICIACIAC	8750
CITACOCGIG ACCCIACAAC CCCCCICGCG AGAGCCGCGI GGGAGACAGC	8800
AAGACACACT CCAGICAATT CCIGGCIAGG CAACATAATC ATGITIGCCC	8850
CCACACTGTG GGCGAGGATG ATACTGATGA CCCATTTCTT TAGCGTCCTC	8900
ATACCACCE ATCACCITGA ACACCICIT AACIGICAGA TCTACCGACC	8950
CIGCIACICC ATAGAACCAC IGGATCIACC TOCAATCAIT CAAAGACICC	9000

FIG. 16E

ATGGCCTCAG	CGCATTTTCA	CTCCACAGTT	ACTOTOCAGG	TGAAATCAAT	9050
AGGGTGGCCG					9100
GAGACACCGG	CCCCCGAGCG	TCCGCCCTAG	GCTTCTGTCC	AGAGGAGGCA	9150
CONTROLLAT	ATGIGGCAAG	TACCICTICA	ACTGGGCAGT	AAGAACAAAG	9200
CTCAAACTCA	CICCAATAGC	GCCCCCIGCC	CCCCICCACT	TGICCGGIIG	9250
GUICACGGCI	GGCTACAGCG	GGGGAGACAT	TPATCACAGC	GIGICICAIG	9300
magaaag	CIGGITCIGG	TTTTGCCTAC	TCCTGCTCGC	TGCAGGGGIA	9350
GCATCTACC	TCCTCCCCAA	CCGATGAAGG	TIGGGGIAAA	CACTCCGGCC	9400
TCTTAAGCCA	TTTCCTGTTT	TITTITITI	TITITITI	TTITICITIT	9450
THEFT	TCCTTTCCTT	CITTITITCC	TITCITITIC	CCTTCTTTAA	950 0
TGGIGGCICC	ATCTTAGCCC	TAGTCACGGC	TAGCIGIGAA	AGGICCGIGA	9550
CCCCCATGAC	TGCAGAGAGT	GCIGATACIG	CCCICICICC	AGAICAIGT	9599

FIG. 16F

10 20 30 40 50 M234567890 1234567890 1234567890 1234567890 1234567890 M234567890 1234567890 1234567890 1234567890 1234567890 M234567890 1234567890 1234567890 1234567890 M234567890 PROPERTY PERAWAGE YEAVALER GELGWART 50 KASERSOFFS REQFIFKARR PERAWAGE YEWLYCHE GELGWART 150 RGSRPSMEFT DERRRENLG KVIDILINGE ADLINGTHEN GARLGGARR 150 LAHGWRLED GWNYTGILP GCSFSIFILA LLSCLTIPAS AYEVRIVSGI 200 MYNINCSNS SIVYEADNI MHTRCYPCV QERISSROW ALIPTILARN 250 ASVPITTIRR HVDLLWGIAA FCSAMYVEL GGSIFLVSQL FIFSPERHET 300 WCDCNISTY GEWSEHRAW DAMMANSPIT ALWSCLIRI FQAWLMAG 350 ALMGVLAGIA YYSMCHWAK VLLVALLFAG VDGEHRIGE VACHITSGFT 400 SIFSSCASK IQUMINGSW HINRTAINON DELQTGFFAA LFYAHKINSS 450 GCPERMASCR PILWFAQGWG PITYIKINSS DQRPYCMYA FRPGGWAS 500 OUGGEVCFT PSEWWATID RSWPYYSWG ENEIDMILIN NIRPPCAMF 550 GCIWMISTGF TRICGFPCN IGGVCRIFLI CPILCFRHP EATYKGASG 600 FWILIPRCLUD YPYRIMYPC TIMESIFKVR MYVGGVERL NAACHWIRGE 650 ROLLEDRIS ELSPLLISTT EWQLLFCAFT TLPALSTGLI HLKANIVDVQ 700 YLWAGVGSAFV SFAIRWEYIL LLFILLADAR VCACLMMIL IAQAFAALEN 750 LWINAASVA GAHGLEFLV FCCAWIKG RLAFGAAVAF YGWPLLILL 800 LALPFRAYAL DITEVAASCGG VVLVELMALT LSPYYKRYIS WCMWLQYFL 850 PWARROQEI LLGPADGWS KGWRLLAPIT AYAQQTRGLI GGITTSLIGE 1000 PVSARROQEI LLGPADGWS KGWRLLAPIT AYAQQTRGLI GGITTSLIGE 1050 KNDJVEGEN IVSTANITEL ATCINGACHT VYHAGGINTI ASPRGEVUQM 1100 PVSARROQEI LLGPADGWS KGWRLAPIT AYAQQTRGLI GGITTSLIGE 1050 KNDJVEGEN IVSTANITEL ATCINGACHT VYHAGGINTI ASPRGEVUQM 1100 PVSARROQEI LLGPADGWS KGWRLAPIT AYAQQTRGLI GGITTSLIGE 1050 KNDJVEGEN IVSTANITEL ATCINGACHT VYHAGGINTI ASPRGEVUQM 1100 PVSARROQEI LLGPADGWS KGWRLAPITA AYAQQTRGLI GGITTSLIGE 1050 KNDJVEGEN IVSTANITEL ATCINGACHT VYHAGGINTI ASPRGEVUQM 1100 PVSARROQEI LLGPADGWS KGWRLAPITA PROCKESTIC PAATAAQGYK 1250 LGTTMRSPVF TINSSPPAVP QSFQAHLHA PIGSKKSINV PAATAAQGYK 1250 PSMDQMKC LITHKPIHA PIPLIMIGA VYNEVILLER TINYMINGE 1550 PGSMUNGKLE		
MSTNPKPORK TKRNINERPO DVKFPGGQI VGSVYLLFRE GPLGVRATE KASERGOFRE REQFIFKARE PERAWAGE YFWPLYCNEE LGWAGNILSP 100 RESRPSWEPT DPRRENING KVIDITICEF ADIMSTIFUV GAPLGCARA 150 LAHGVRVLED GNYATGNIP GCSFSIFILA LISCLITIPAS AVEVRNYGGI 200 YHVINICSNE SIVYEAADVI MHTPGCVPCV QEINSSROW ALIPTILARN 250 ASVPITTIRE HVILLIUGIAA FCSAMYUGIL CGSIFLINGIL FIFSPRRHET 300 VQDCNCSITY GHVSEHMAW IMMINSPIT ALWOQLIEI PQAVUMMAG 350 AHAGVIAGIA YYSMANMAK VLIVALIFAG VUCEHHTIGE VACHTISGFT 400 SIFSSASQK IQLWINGSW HINRIAINEN DELQTGFFAA LFYAHKFISS 450 GCPERMASCR PILWFAQGW PITYTKRINSS DEPFUCMYA PRECGWPAS GCIWMISTEF TKICGGFEN IGGVENEILI CPILCFKHP EATYTKCGSG GNLEDRISS ELSFLLISTT EWILLFCAFT TLPALSTRLI HHENTIVIDVQ 700 YLXGVGSAFV SFAIKWEYIL LIFLLIADAR VUCALIMMLI NARACWIRGE GSO ROLLEDRISS ELSFLLISTT EWILLFCAFT TLPALSTRLI HHENTIVIDVQ 701 YLXGVGSAFV SFAIKWEYIL LIFLLIADAR VUCALIMMLI LAQAEAALEN 750 ILQAFLAVP YFVRVQGLIR ICALARKIAG GHVUPATIK LIALIFGFUW 900 ILQAFLAVP YFVRVQGLIR ICALARKIAG GHVUPATIK LIALIFGFUW 901 ILQAFLAVP YFVRVQGLIR ICALARKIAG GHVUPATIK LIALIFGFUW 902 ILGTIMRSVF TINSSPRAVP OSFOVAHLAFT AYAQQTRGLI GCITTSLITGR 1050 PKRQVEGEVQ IVSTATQTFL ATCINGVCH VYHAGGTRII ASPKOPVIQM VTNVDQDIMG WPARGGSRSL TPCTGGSSLL YLVTRHADUT PVRRREDSRG 1L150 PRESVIVSHP NIEEVALSTI GEIFFYGRAI FRAVUTRIGE PUTSTYGGEFL 1150 VLVINPSVAA TLGFGAYMSK AHGVDRIFT GYPTTTIGSP TITYSTYGGEFL 1300 LGTIMRSVF TINSSPRAVP OSFOVAHLAF PIGSCKSTKV PAAVAAQGYK 1250 VLVINPSVAA TLGFGAYMSK AHGVDRIFT GYPTTTIGSP TITYSTYGGEFL 1300 VLDINICVIQ TVDFSLDPTF TIETTTLQD AVSRTORRE TERGERGITR 1500 PERSVIVSHP NIEEVALSTI GEIFFYGRAI PIGSCKSTKV PAAVAAQGYK 1250 VLDINICVIQ TVDFSLDPTF TIETTTLQD AVSRTORRE TERGERGITR 1500 PERSVIVSHP NIEEVALSTI GEIFFYGRAI PLEVIKGGER TERGERGITR 1500 PERSVICHUR VVFGSLOHILA HTLSQTKQSG ENEFLYNIXQ 1450 PERSWILMI ATTURING VVFGCEWILL PETTVIRIC AYMITGEFLY 1600 PERSWILMI ATTURING VVFGCEWILL PETTVALTA AYMI		
RASERSOFIC RROPIEKARR PECRAWAQCE YEWFLYCNEG LGWAGNILSP RGSRPSMIPT DERRRSRIG KVIDILITOGF ADLMSYIFLV GAFLGGARA 150 IAHSVRVLED GMYATCHLP CCSFSIFLIA LLSCLTIPAS AYEVRWSGI 200 YHVINDCSIS SIVYEAADMI MHIRCCYPCV QENISSROW ALITPILAARN ASVPITITIR HVDLLWGIAA FCSAMWEDL CGSIFLVSQL FTFSPRHET 300 ASVPITITRR HVDLLWGIAA FCSAMWEDL CGSIFLVSQL FTFSPRHET 300 VQDCNCSIYP GHVSCHRAW INMAMSPIT ALWOQLIRI PQAVVIMAG 350 AHASALAGIA YYSMCHWAR VLIVVALLFAG VDEHTHTIG VACHITSEFT 400 SLFSSCASQK IQLWININSW HINRIAINON DSLQTGFFAA LFYARKFINSS 550 GCPERMASCR PIDWFAQGAI PITYIKRINS DQRFYCMYA FRECGAVPAS 500 QVCGPVYCFT PSPVWGIID RSGVFTYSW ENEIDMILIN NIRPPQGWF 550 GCIWANSIGF TKTCGSPEN IGSVGWRILI CPIDCFRRH EATYIKGGG 600 GWLIPPRCLVD YPYRLWHYP TIMFSIFKVR MYVGSVEHL NAACMIRGE 650 RCHEIRRS ELSPILLSTI ENQILECAFT TIPALSTGILI HLQNIVUNQ 700 YLYGYGSAFV SFAIRWEYIL LIFILLADAR VCACLUMILI IAQAEAALEN 750 LALPFRAYAL DTEVAASCGG VVIWELMALT LSPYKRYIS WLMWLLYFL 850 ILQASLLKUP YPVRVGGLIR ICALARKIAG GHYVQMAIK LLALIGIPUV 950 NHLTPLRWA HNSIRDLAVA VEPVVFSRME TKLTTWADIT AACCDIINEL 1000 PVSARROQEI LLGPACGWS KGWILLACUT HPILVFDITK LLALIGPIW 950 NHLTPLRWA HNSIRDLAVA VEPVVFSRME TKLTTWADIT AACCDIINEL 1000 PVSARROQEI LLGPACGWS KGWILLACUT VHYBAGIRIT ASPKGFVIQM 1150 PVSARROQEI LLGPACGWS KGWILLACUT VHYBAGIRIT ASPKGFVIQM 1150 PVSARROQEI LLGPACGWS KGWILLACUT FAAVVITAGAI ASPKGFVIQM 1150 PVSARROQEI LLGPACAWS AHGDINIT GVTTTIGSP TYSTKEFL 1300 PVSARROQEI LLGPACAMS AHGDINIT GVTTTIGSP TYSTKEFL 1300 PROSVITAGH NIEWALSTI GEIFFYGRAI PLEVVIQGRH LIFCHSKKC 1400 PROSVITAGH PROSVICEC YAGCAWEL THETVIVAR AYMTPELPV 1500 PPSALQWKC LIFIKPTIHG P		
RESRESMET DEPRRESNIG KVIDILITOEF ADIMSTIFLY GAPLOGARA LAHSVRVLED GVNYATCHLP CCSFSIFILA LISCLITIBAS AYEVRNYSSI 200 YHVINDCSIS SIVYEAADVI MHTPGCVCV QEINSROW ALITYLAARN ASVETTITIR HVDLIMSIAA FCSAMYWEDL CGSIFLYSQL FIFSPRRHET 300 WODNCSTYP GHVSHRWW DMMWSPIT ALWYQLIRI PQAVVIMMAG 350 AHWSVLACIA YYSMWAWAK VLIVALIFAG VDGETHITER VACHITSGFT 400 SLFSGASQK IQLWINGSW HINRTAINCN DSLQTGEFAA LFYAHKFISS 450 GCPERMASCR PIDWFAQOMS PITTYKENSS DQRFYCMYA PRECGVVPAS QVCGPVYCFT PSFWWITD RSSVPIYSWS ENEDMILN NIRPPQSWF 500 GCIMANSIGF TKICGSPEN IGGVENRILL CPIDCFRHP EATYTKOGSG GCOMMISTIGF TKICGSPEN IGGVENRILL CPIDCFRHP EATYTKOGSG GCOMMISTIGH TRANSPORT		
NATIONAL DE GENYATCHE CESFSIFILA LISCLITEAS AYEVRANGI 250 YHVINDESES SIVYEAADVI MHIRECAPEV QEENSERGAN ALITEIAARN 250 ASVPITTIRR HUDILWGIAA FCSAMWEDI CESIFLVSQL FIFSPRRHET 350 WQDCNETYP GHVSHRMAN IMMISPIT ALWOQLIRI PQAVVIMAG 350 HAUSTLAGIA YYSMCHARK VILIVALIFAG VDSETHITER VACHITSSFT 400 SLFSGASQK IQLMINGSM HINRIAINEN DSLQTGFFAA LFYAHKINSS 450 GCPERMASCR PIDWFAQGAS PITYMENSS DQRPYCHYA PRPCGAVPAS 500 QVGGPVYCFT PSPAVVGIID RSGAVRILLI CPIDLFRKHP EATYIKGSG 600 QVGGPVYCFT PSPAVVGIID RSGAVRILLI CPIDLFRKHP EATYIKGSG 600 QVLGPVXGAT PSPAVVGIIT EAQILPCAFT TLPALSIGLI HLIQNIVIAVQ 700 ROLLEDRISS ELSPILLSTT EAQILPCAFT TLPALSIGLI HLIQNIVIAVQ 700 YLYGVGSAFV SFAIKMEYH. LIFILLADAR VCACLIMMIL IAQABAALEN 750 LVLUNAASVA GAHGILSFLV FFCAAWYKG KAPGAAVAF YGWPLILLI 850 LALPPRAVAL DTEVAASCGG VALMEMALT LSPYKKYTS WLMWILQYFL 850 TRVEQQLHW VPPINAGGF DAVILINCAV HPILVFDITK LLLAIFGPIW 900 HLITELRIAA HUSHRILAVA VEFAVFSRME TKLITWGADI AACGDIINGL 1000 PVSARRQGEI LIGPADGAMVS KAPKALAPIT AYFQQIRGIL GCIITSLIGEN 1000 PVSARRQGEI LIGPADGAMVS KAPKALAPIT AYFQQIRGIL GCIITSLIGEN 1100 PVSARRQGEI LIGPADGAMVS KAPKALAPIT AYFQQIRGIL GCIITSLIGEN 1150 PKNQVESSOQ PILICDECHS TICTITGSCANI VARGAGRINI ASFKGEVIQM 1100 PVSARRQGEI LIGPADGAMVS KAPKALAPIT GARTHITGSP ITYSTYCKFL 1300 LGTIMPSEVF TINSSPPAVP QSFQVAHLAP PIGSEKSIKV PAAYAAQGYK 1250 PROSVIVSHP NIESVALSTI GEIPFYGKAI PLEVIKGGRH LIFCHSKKC 1400 PROSVIVSHP NIESVALSTI GEIPFYGKAI PLEVIKGGRH LIFCHSKKC 1400 PPSADQAMKC LIRIKPTIHG PIPLYRIGA VQAEVILIAP TIXYIMICMS 1650 PPSADQAMKC LIRIKPTIHG PIPLYRIGA VQAEVILIAP TIXYIMICMS 1650 PPSADQAMKC LIRIKPTI		
ANTINICSNS SIVYEAADVI MHIPCOVCV QEINSSROW ALITHIAARN 250 ASVPITTIRR HVOLLUGIAA FOSAMYCEL COSIFINSOL FIRSPRIET 300 WODONSTYP GHVSHRMAW IMMMSPIT ALWSOLIRI PQAVVINAG 350 AHGVLAGIA YYSMOMAK VLIVALIFAG VDEHHITCR VACHITSGIT 400 SLFSGASK IQLMINISSW HINRIANN DELQTIGFFAA IFYAHKINSS 450 GCPERMASCR PILWFAQGWG PITYIKHNSS DQRPYCMYA PRPCGVVPAS 500 QVCGPVYCFT PSPWWGITD RSGVPIYSWG ENEDMILN NIRPPQGWF 550 GCTWMNSTGF TKTCGGPFON IGSVENRILI CPITCFRRIP EATYIKCGSG 600 EWLIPRCIND YPYRUMYPC TINFSIFKUR MYVGSVEHRI NACMIRCE 650 ROLLEDROS ELSPILLSTI ENQILPCAFT TLPALSTGLI HLKNIVIVQ 700 YLYGGGSAFV SFAIKWEYIL ILIFLILADAR VACCIUMMIL IAQAFAALEN 750 LWINAASVA GAHGILSFILV FFCAAWYKG RLAPGAAYAF YGWPILILL 800 LALPPRAYAL DTEVAASOG WIMEMALT LSPYYRRYIS WIMMLQYFL 850 TRVEAQLHW VPPINVRGGR DAVILIMCVV HPTLVFDITK LLIALFGPLW 900 HILTPIRWA HNGLRDLAVA VEFWYSRME TKLITIVADIT AACCDIINGL 1000 PVSARRQOEI LIGGADGWS KGWRLAPIT AYAQQTGGL GCITTSLTGR 1050 DKNQVESEVQ IVSTATQTFL ATCINGVCWI WHAGGIRTI ASPKGPVIQM 1100 YTMNDQDLW WPAPGGSRSL TPCTCGSSDL YLVTRIADVI PVRREDEGG 1150 SLLSPRISY LKGSSGGFLL CACHANCAL PITT AYAQQTGGL GCITTSLTGR 1050 ADCCCSGGAY DILICDECHS TDATSLIGIG TVLDQAETAG ARLWLATAT 1350 PRGSVIVSHP NIEEVALSTT GEIFFYGRAI PIGSEKSTKV PAAYAAQGK 1250 VIJUNIPSVAA TLGFGAYMKK AHGJDRINT GWRTTTIGSP ITYSTYCKFL 1300 DELAAKIVAL GINAVAYRG LDVSVIPTSG DVVSKTARGR TERGKGTYR 1450 VIJUNIPSVAA TLGFGAYMK AHGJDRINT GWRTTTIGSP ITYSTYCKFL 1300 PRGSVIVSHP NIEEVALSTT GEIFFYGRAI PLEVIKGGRH LIFCHSKKKC 1400 PPGSVIVSHP NIEEVALSTT GEIFFYGRAI PLEVIKGGRH LIFCHSKKKC 1400 PPGSVIVSHP NIEEVALSTT GEIFFYGRAI PLEVIKGGRH LIFCHSKKKC 1400 PPGSVIVSHP NIEEVALSTT GEIFFYGRAI PLEVIKGGRH LIFCHSKKKC 1450 PPGSVIVSHP NIEEVALSTT GEIFFYGRAI PLEVIKGGRH LIFCHSKKKC 1400 PPSMDQMKC LIRLKPILMG PTPLLYRIGA VQUENTIHP TIKYIMICMS 1650 ADLEWTSTW VLVGGVIAAL AAYCLSTGCV VTVCRIVING KPALIPRAV 1750 LYCGFDEMEE CSQHLPYIEQ GMLAPGFKQ KALGLIQTAS RHAEVITPAV 1750 LYTGGTBEME CSCHLPYIEQ GMLAPGFKQ KALGLIQTAS RHAEVITPAV 1750 LYTGGTBEME CSCHLPYIEQ GMLAPGFKQ KALGLIQTAS RHAEVITPAV 1750 LYTGGTBEME CSCHLPYIEQ GMLAPGFKQ KALGLIQTAS SVGLGKVUND 1850		
ASVPTTTIRR HVDLLVSIAA FCSAMYVEL CGSIFLYSQL FIFSPRRHET VQDCNCSIYP GHYSCHRAW IMMANSPIT ALWYSQLIRI PQAVVIMAG SSO AHMANIACIA YYSMCAMAK VLIVALIFAG VOCETHITER VACHITISET 400 SLFSSGASK IQLWINISSW HINRIALNIN DSLQTUFFAA LFYAHKFISS SCPERMASCR PILWFAQAG PITYIKANSS DQRPYCMHA PRRCGWPAS GCYCGPYYCFT PSPVWUTID RSGVPTYSW ENEIDMILN NIRPPQAWF GCTWMSTGF TKTCGGPFON IGGVCRITLI CPIDCFRKIP EATYTKCSS GCTWMSTGF TKTCGGPFON IGGVCRITLI CPIDCFRKIP EATYTKCSS GCUMITERCIND YPYRIMYPC TIAFSIFKVR MYVGSVEHIL NAACMIRGE FSO RCLIEDROS ELSFILLSTI EWQILPCAFT TLPALSTGLI HLHQNIVIVQ 700 YLYGVGSAFV SFALWEYLI LIFLIADAR VCACLUMMIL TAQAFAALBN 750 LWINAASVA GARGILSFLV FFCAAWYTKG RLAPGAAYAF YGWPLLILL BOO LALPPRAYAL DTEVAASCGG VLIVEIMALT LSPYYKRYIS WCMWLQYFL STO TRVEAQLHW VPPINVRGER DAVILINGVU HPTILVFDITK LLLAIFGPIW 900 HLIQASLLKUP YFVRVQELIR ICALARKIAG GHYVCMAIK LGALTGTVVY 950 NHLTPIRDWA HNEIDLAVA VEFVYFSRME TKLLTTUGADI AACCDIINGL DKNQVESEVQ IVSTATQTFL ATCINGVCMT VYHCAGIRTI ASPKGPVIQM 1100 PYSARROQEI LLGEADGWYS KGWRLLAPIT AYAQQTRGLL GCITTSLTGR DKNQVESEVQ IVSTATQTFL ATCINGVCMT VYHCAGIRTI ASPKGPVIQM 1100 YINVDQDLWG WPAPQGSRSL TECTOGSSDL YLMTRIADVI PVRRREDERG SLISPRISY LKGSSGGFLL CPACHAVLET RAAVCTRGVA KAVDFIPVEN LGTTMFSFVF TINISSPPAVP QSFQVAHLHA PTGSCKSTKV PAAYAAQGYK LUCINGVSAP TILGEGANMEK AHGYDPINET GYRTTTIGSP ITYSTYCKFL ADGCCSGGAY DIIICDECHS TDATSLIGIG TVLDQAETAG ARLWLATAT ADGCCSGGAY DIIICDECHS TDATSLIGIG TVLDQAETAG ARLWLATAT ADGCCSGGAY DIIICDECHS TDATSLIGIG TVLDQAETAG ARLWLATAT 1350 PRGSVIVSHP NIEEVALSTT GEIFFYGKAI PLEVIKGGRH LIFCHSKKKC 1400 PPSWDQMKC LIRLKPTLHG BYBLYRIGA VQUENTINEP TIRKYIMICMS ADLEWTSIW VLNGGVLAAL AAYCLSTGCV VIVAENUSG RPALIPREV 1750 COPHEFFOR VFTELIHIDA HFLSQTKQSG ENFFYLVAYQ ATVCARQAP 1600 PPSWDQMKC LIRLKPTLHG GYMLAEGEKG VURSELLIGTAS RHAEVITPAV 1750 LUTTGOTLLIN ILGSWAAQL AAYGAATAFV GAGLAGAIG SVGLGKVIVD 1850 LITTGOTLLIN ILGSWAAQL AAYGAATAFV GAGLAGAIG SVGLGKVIVD		
VODCNSTYP GHVSHRMW IMMINSPIT ALWSCLEI PQAWIMVAG AHMSVLAGIA YYSMCHMAK VLIVALLIFAG VDGEIHTIGR VACHITSGFT 400 SLFSCASQK IQLMINGSW HINRIAINCN DSLQTGFFAA LFYAHKINSS 450 GCPERMASCR PIDWFAQGAG PITYIKKINSS DQRPYCMYAP RRCGWPAS 500 QVCGFVYCFT PSPWYGITD RSGVPYYSW ENEIDMILIN MIRPPQSWF 550 GCTMMSTGF TKTCGGPPCN IGGWARTILI CPIDCFRKHP EATYIKGSG 600 FWLITERCIMD YPYRIMHYC TIMFSLFKUR MYVGGVEHRL NAACHWIRGE 650 RCNLEDRDS ELSPILLSTT EWQILPCAFT TLPALSTGLI HLHQNIVIMQ 700 YLYGVGSAFV SFAIKWEYIL LIFLILADAR VCACLMMILI TAQAFAALEN 750 IMMINASVA GAHGILSFLV FFCAAWYIKG RLAPGAAYAF YGWFILLILL 800 LALPPRAYAL DTEVAASCG VUWGIMALT LSPYYKRYIS WCMWILQYFL 850 TRVEAQLHW VPPINVRGGR DAVILLMCVV HPILVFDITK LLILAIRGFUW 900 HILQASLIKUP YFVRVQGILR ICALARKIAG GHYVQMAIIK LGALTGTYVY 950 NHLITPLRIMA HNGLRDLAVA VEPVFSRME TKLITWGADT AACCDIINGL 1000 PVSARRQQEI LLGPADGWS KGWRLLAPIT AYAQQTRGLL GCITTSLITGR 1050 DKNQVEGENQ INSTATQTFL ATCHNGVOT VYHAGARTHI ASFKGFVIQM 1100 YINMDQDUG WPAPQGSRSL TPCTCGSSUL YLVTRHADVI PVRRCEDSRG 1150 SILSPRPISY LKGSSGGFLL CPACHAVGLF RAAVCTRGVA KANDFIFVEN 1200 LGTIMRSFVF TINSSPPAMP QSFQVAHLHA PTGSGKSIKV PAAYAAQGYK 1250 VLVINPSVAA TLGFGAYMSK AHGVDENIRT GVRTTITGSP HTYSTYGKFL 1300 ADGCCSGAY DILICDECHS TDATSLIGIG TVLDQAETGG ARIAVLATAT 1350 LGTIMRSFVF TINSSPPAMP QSFQVAHLHA PTGSGKSIKV PAAYAAQGYK 1250 VLVINPSVAA TLGFGAYMSK AHGVDENIRT GVRTTITGSP HTYSTYGKFL 1300 FVARGERPSG MFDSSVLCEC YDACCAWYEL TPAETIVRIR AYMNTRGLFV 1550 VIDCNICVIQ TVDFSLDPTF TIETTTLQD AVSKTORRGR TERERGGTYR 1500 FVARGERPSG MFDSSVLCEC YDACCAWYEL TPAETIVRIR AYMNTRGLFV 1550 FVARGERPSG MFDSSVLCEC YDACCAWYEL TPAETIVRIR AYMNTRGLFV 1550 PPSWDQMKC LIRKPTHG PTPILLYRIGA VQNEVILITE TIKYIMICMS 1650 PPSWDQMKC LIRKPTHG PTPILLYRIGA VQNEVILITE TIKYIMICMS 1650 ADLEWYISTW WLVGGWIAAL AAYCLSTGCV VVVCRIVLES RAHIPDREV 1700 LYTGOTLIFN ILGGWAAQL AARGAATAFV GAGLAGAAIG SVELEKVLVID 1850		
AHWVIAGIA YYSMCHWAK VLIVALIFAG VDGEIHTIGR VACHITSGFT SLFSSCASOK IQLWNINGSW HINRIAINCN DSLQTGFFAA LFYAHKFNSS GCPERMASCR PIDWFAQAG PITYIKANSS DORPYCHYA PRPCGVVPAS GCOGRVYCFT PSRVWGTID RSGVFTYSW ENEIDWALIN NIRPPQGWF SSO GCGEVYCFT PSRVWGTID RSGVFTYSW ENEIDWALIN NIRPPQGWF SSO GCTGMNSTGF TKTCGGPPCN IGGVGRILL CPIDCFKHP EATYTKCGSG GCOMMITTERCUP YPYRLWHYPC TLNFSIFKUR MYVGGVEHIL NAACMURGE SSO RULTPRCLUD YPYRLWHYPC TLNFSIFKUR MYVGGVEHIL NAACMURGE SULVINAASVA GAHGILSFLV FFCAAWYIKG RLAFGAAYAF YGWPLLILL 800 LALPPRAYAL DIEVAASCGG VVIVEIMALT LSPYYRRYTS WLMWLQYFL 850 TRVEAQLHW VPPINNRGGR DAVILLMCVV HPILVFDITK LLIAIFGPLW 900 LIQASLLKVP YFVRVCGLIR ICALARKIAG GHYVQMAILK LGALTGTYVY 950 LIQASLLKVP YFVRVCGLIR ICALARKIAG GHYVQMAILK LGALTGTYVY 950 PVSARRQQEI LLGPADGWS KGMRLLAPIT AYAQQTRGLL GCTITTSLIGER 1050 DKNQVEGENQ IVSTATQTFL ATCINGVCWI VYHQAGTRII ASPKGFVIQM 1100 YTMVDQDLW WPAPQGSRSL TPCTCGSSDL YLVTRHADVI PVRRRCDSRG 1150 SLLSPRPISY LKGSSGGFLL CPACHAVGLF RAAVCTRGJA KAVDFITVEN 1200 LGTIMRSPVF TINSSPPAVP QSFQVAHLHA PTGSGKSTKV PAAYAAQGYK 1250 VIDANDSVAA TLGFGAMSK AHSVDENIRT GVFTTTGSP TYYSTYGKFL 1300 LGTIMRSPVF TINSSPPAVP QSFQVAHLHA PTGSGKSTKV PAAYAAQGYK 1250 VIDANDSVAA TLGFGAMSK AHSVDENIRT GVFTTTGSP TYYSTYGKFL 1300 LGTIMRSPVF TINSSPPAVP QSFQVAHLHA PTGSGKSTKV PAAYAAQGYK 1250 VIDANLAVAL GINAVAYYRG LDVSVIPTSG DVVVVSTDAL MTGFTCDFDS VIDANCKLEVE MFLEVILLG PTPLYRICA VANTPGLEV TASTILIGI TVLDQAETAG ARLWLATAT 1350 PRGSVIVSHP NIEEVALSTT GEIFFYGKAI PLEVIKGGH LIFCHSKKC 1400 PPSNDQMKC LIRLKPTILM FTPLLYRIGA VQNEVTLIRP TTRYTMICMS ADLEVVTSW VLUGGVIAAL AAYCLSTGCV VIVETVLSG KPALTPREV 1550 ADLEVVTSW VLUGGVIAAL AAYCLSTGCV VIVETVLSG KPALTPREV 1750 LYQEFDMEE CSQHLFYIEQ GMLAEDFKQ KALGLIQTAS RHAEVITPAV 1750 LTTGOTLIFN 11GGWAAQL AARGAATAFV GAGLAGAAIG SVELGKUUND 1850		
SIFSCASOK IQLMINISSW HINRTAIN'N DSLQTGFFAA LFYAHKFNSS 500 GCPERMASCR PILWFAQGWS PITYTKENSS DQRPYCWHYA PRPCGWPAS 500 QWGGPYYCFT PSPWWGTID RSGYPTYSWG ENEITDMILIN NIRPPQGWF 550 GCTWMNSTGF TKTCGGFPCN IGGVANRILL CPIDCFRKHP EATYTKCGSG 600 PWITPRCLVD YPYRLMYPC TINFSIFKVR MYVGGVERL NAACMWIRGE 650 RCNLEDRDRS ELSPLLISTT EWQILPCAFT TLPALSTGLI HHANIVDVQ 700 YLYGVGSAFV SFAIKWEYHL LLFLLIADAR VCACLMMIL IAQAFAALEN 750 LWINAASVA GAHGILSFLV FFCAAWYKG RLAPGAAYAF YGWPLLILL 800 LALPPRAYAL DTEVAASCGG VULVELMALT LSPYYKRYIS WCMWLQYFL 850 TRVEAQLHW VPPINVRGGR DAVILLMCVV HPTLVFDTIK LLIAIFGPLW 900 HLQASLLKVP YFVRVQELIR ICALARKTAG GHYVQMAILK LGALIGTYVY 950 NHLITPLRWA HNYERDLAVA VEPVVFSRME TKLITWGADI AACCDIINGL 1000 PVSARRQQEI LLGPADGWS KGWRLIAPIT AYAQOTRGLI GCIITSLITGR 1050 DKNQVEGEVQ IVSTANDIFL ATCINGVCMI VYHCAGIRTI ASPRGPVIQM 1100 YINVDQDLUG WPARGGSRSL TPCTCGSSDL YLVTRHADVI FVRRGDSRG 1150 SLLSPRPISY LKGSSGGFLL CPACHAVGLF RAAVCTRGVA KAVDFIPVEN 1200 LGTIMRSFVF TINSSPPAVP QSFQVAHLHA PIGSGKSTKV PAAYAQGYK 1250 VLIVINPSVAA TIGGGAYMSK AHSVDRNIRT GVRTITTGSP TYSTYGKFL 1300 ADGCCSGGAY DILICDECHS TDATSILGIG TUDQAETAG ARLWLANTAT 1350 PRGSVIVSHP NIEEVALSIT GEIFFYGKAI PLEVIKGGRH LIFCHSKKKC 1400 DELAAKIMAL GINAVAYRG LIDSVIPTSG DVVVVSTDAL MIGFTCDFDS 1450 VIDONICVIQ TVDFSLDPIF TIETITLPQD AVSRTQRGR TERGKPGIYR 1500 PPSWDQMKC LIRLKPTIMS PIPLLYRIGA VQNEVILIHP TIKYIMICMS 1650 ADLEWTSTW VLVGGVAAL AAYCLSTGCV VIVERIVLSG KPALIPLEPU 1750 ADLEWTSTW VLVGGVAAL AAYCLSTGCV VIVERIVLSG KPALIPLEPU 1750 QTIMQKLEVF WAKHMIFTS GIQYLACLST LPCNPATASL MFTAAVTSP 1800 LITIGOTILFN ILGSWAAQL AAPGAATAFV GAGLAGAAIG SVELGKVIVD		350
QCPERMASCR PIDWFAQGMS PITYIKHNES DQRPYCMHYA PRPCGWPAS 500 QVCGPVYCFT PSPVWGITD RSGVPTYSMS ENETDMILIN NIRPPQGMWF 550 GCIMMISTEF TKTCGGPPCN IGGVCNRTLI CPIDCFRKHP EATYTKCGSG 600 HMLTPRCLUD YPYRIMHYPC TINFSIFKVR MYCGGVERL NAACNWIRGE 650 RCNLERRES ELSPLLISTT EWILLPCAFT TLPALSTGLI NAACNWIRGE 650 YLYGVGSAFV SFALKWEYTL LLFILLADAR VCACLMMIL IAQAEAALEN 750 LLVINAASVA GAHGILSFLV FFCAAWYIKG RLAPGAAVAF YGMPILLIL 800 LALPPRAYAL DTEVAASCGG VVLVEIMALT LSPYYKRYIS WLMWLQYFL 850 TRVEAQLHW VPPINVRGGR DAVILLMCVV HPILVFDITK LLLAIFGPLW 900 HLIPLRDWA HNGIRDLAVA VEPVVFSRME TKLITWGADT AACCDIINGL 1000 PVSARRQQEI LLGPADGMVS KGWRLIAPIT AYAQOTRGIL GCITTSLITGR 1050 DKNQVEGEVQ IVSTATQTFL ATCINGVCMT VYHAAGHRIT ASPKGPVIQM 1100 YINVLODLUG WPAPQGSRSL TPCTCGSSDL YLVTRHADVI PVRRREDSRG 1150 SLLSPRPISY LKGSSGPLL CPACHAVELF RAAVCTRGVA KAVDFIPVEN 1200 LGTIMRSPVF TINSSPPAVP QSFQVAHLHA PIGSGKSTKV PAAYAAQGYK 1250 VLVINPSVAA TLGFGAMSKA AKSVDPNIRT GVRTTTIGSP ITYSTYGKFL 1300 ADGCCSGGAY DILICIDCHS TDATSLIGIG TVLDQAETAG ARLWLATAT 1350 PRGSVIVSHP NIEEVALSTT GEIPFYGKAI PLEVIKGGRH LIFCHSKKKC 1400 PPGSVIVSHP VVDFSLDPTF TIETTILPQD AVSRTQRGR TGRGKFGTYR 1500 FVARGERPSG MFDSSVLCCC YDAGCAWYEL TPAETIVRIA AYMITELIPV 1550 CQCHLEFWEG VFTGLIHHIA HFLSQIKGGG ENFYLVAYQ AIVCARAQAP 1600 PPSWDQMKK LIRLKPILIG PTPLLYRIGA VQNEVILIHP TIKYIMCMS 1650 ADLEVYSTW VLVGGVLALA AAYCLSTOCV VVVGRIVLSG KPALIPDREV 1700 LYTQGFDEMEE CSQHLPYIEQ CMMLAEQFKQ KALGILQTAS RHAEVITFAV 1750 LYTQGTLLFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVELCKVIVD 1850	AHWGVLAGLA YYSMWGWWAK VLIVALLFAG VDGETHITGR VACHTISGFT	400
QVCGPVYCFT PSPWWGTID RSGVPTYSWG ENEIDMILN NIRPPQCNWF 550 GCIWNSTGF TKTCGSPPCN IGGVCNRTLI CPIDCFRKHP EATYTKCSSG 600 HWLTPRCLVD YPYRLWHYPC TLNFSIFKVR MYVGGVEHEL NAACMURGE 650 RCNLEDRIS ELSPLILSTT EMQILPCAFT TLPALSTGLI HLHQNIVDWQ 700 YLYGVGSAFV SFAIKWEYIL LLFILLADAR VCACLMMILL IAQAFAALEN 750 LVVINAASVA GAHGILSFLV FFCAAWYKG RLAFGAAYAF YGWPLLLLL 800 LALPPRAYAL DTEVAASCGG VVLVGLMALT LSPYYKRYIS WCMWLQYFL 850 TRVEAQLHW VPPINNRGGR DAVILLMCVV HPTLVFDITK LLLAIFGPLW 900 HIQASLLKVP YFVRVQGLIR ICALARKIAG GHYVQMAIIK IGALTGTYVY 950 NHLTPLRDWA HNGLRDLAVA VEPVVFSME TKLITWADIT AACGDIINGL 1000 PVSARRQQEI LLGPADGWVS KGWRLLAPIT AYAQQTRGLL GCITTSLTGR 1050 DKNQVEGEVQ IVSTATQTFL ATCINGVCWI VYHCAGTRTI ASFKGPVIQM 1100 YTNNDQDLWG WPAPGSSRSL TPCTCGSSDL YLWTRHADVI PVRRRGDSRG 1150 SLLSERPISY IKGSSGGFLL CPACHAVGLF RAAVCTRGVA KAVDFIPVEN 1200 LGTIMRSPVF TINSSPPAVP QSFQWAHLHA PTGSGKSTKV PAAYAAQGYK 1250 VLVINPSVAA TLCFGAYMSK AHGVDPNIRT GVRTTITGSP TTYSTYGKFL 1300 PRGSVTYSHP NIEEWALSTT GEIPFYGKAI PLEVIKGGRH LIFCHSKKC 1400 PRGSVTYSHP NIEEWALSTT GEIPFYGKAI PLEVIKGGRH LIFCHSKKC 1400 PRGSVTYSHP NIEEWALSTT GEIPFYGKAI PLEVIKGGRH LIFCHSKKC 1400 PPSWDQMKC LIRLKPTLAG TTETTTLPQD AVSRTQRGR TCRCKFGTYR 1500 FVAPGERPSG MFDSSVLCEC YDAGCAWYEL TPAETTVRIR AYMNTFGLEV 1550 CQDHLEFWEG VFTGLIHIDA HFLSQTKQSG ENFYLWAYQ ATVCARAQAP 1600 PPSWDQMKC LIRLKPTLAG PTPLLYRIGAV VQNEVILTHP ITKYTMICMS 1650 ADLEVVTSW VLVGSVLAAL AAYCLSTGCV VTVGRIVLS KPALITPREV 1700 LYQEFDEMEE CSCHLPYTEQ GMLAEDFKQ KALGLIQTAS RHAEVITPAV 1750 QTMWKLEVF WARHMWFTS GIQVLAGLST LPCNPAIASL MAFTAAVTSP 1800 LITTGOTILIFN ILGGWAAQL AAPCLATAFV GAGLAGAAIG SVGLGKVUD 1850	SLFSSGASOK IQLVNINGSW HINRIALNON DSLQIGFFAA LFYAHKFNISS	450
QVCGPVYCFT PSPWWGTID RSGVPTYSWG ENEIDMILN NIRPPQCNWF 550 GCIWNSTGF TKTCGSPPCN IGGVCNRTLI CPIDCFRKHP EATYTKCSSG 600 HWLTPRCLVD YPYRLWHYPC TLNFSIFKVR MYVGGVEHEL NAACMURGE 650 RCNLEDRIS ELSPLILSTT EMQILPCAFT TLPALSTGLI HLHQNIVDWQ 700 YLYGVGSAFV SFAIKWEYIL LLFILLADAR VCACLMMILL IAQAFAALEN 750 LVVINAASVA GAHGILSFLV FFCAAWYKG RLAFGAAYAF YGWPLLLLL 800 LALPPRAYAL DTEVAASCGG VVLVGLMALT LSPYYKRYIS WCMWLQYFL 850 TRVEAQLHW VPPINNRGGR DAVILLMCVV HPTLVFDITK LLLAIFGPLW 900 HIQASLLKVP YFVRVQGLIR ICALARKIAG GHYVQMAIIK IGALTGTYVY 950 NHLTPLRDWA HNGLRDLAVA VEPVVFSME TKLITWADIT AACGDIINGL 1000 PVSARRQQEI LLGPADGWVS KGWRLLAPIT AYAQQTRGLL GCITTSLTGR 1050 DKNQVEGEVQ IVSTATQTFL ATCINGVCWI VYHCAGTRTI ASFKGPVIQM 1100 YTNNDQDLWG WPAPGSSRSL TPCTCGSSDL YLWTRHADVI PVRRRGDSRG 1150 SLLSERPISY IKGSSGGFLL CPACHAVGLF RAAVCTRGVA KAVDFIPVEN 1200 LGTIMRSPVF TINSSPPAVP QSFQWAHLHA PTGSGKSTKV PAAYAAQGYK 1250 VLVINPSVAA TLCFGAYMSK AHGVDPNIRT GVRTTITGSP TTYSTYGKFL 1300 PRGSVTYSHP NIEEWALSTT GEIPFYGKAI PLEVIKGGRH LIFCHSKKC 1400 PRGSVTYSHP NIEEWALSTT GEIPFYGKAI PLEVIKGGRH LIFCHSKKC 1400 PRGSVTYSHP NIEEWALSTT GEIPFYGKAI PLEVIKGGRH LIFCHSKKC 1400 PPSWDQMKC LIRLKPTLAG TTETTTLPQD AVSRTQRGR TCRCKFGTYR 1500 FVAPGERPSG MFDSSVLCEC YDAGCAWYEL TPAETTVRIR AYMNTFGLEV 1550 CQDHLEFWEG VFTGLIHIDA HFLSQTKQSG ENFYLWAYQ ATVCARAQAP 1600 PPSWDQMKC LIRLKPTLAG PTPLLYRIGAV VQNEVILTHP ITKYTMICMS 1650 ADLEVVTSW VLVGSVLAAL AAYCLSTGCV VTVGRIVLS KPALITPREV 1700 LYQEFDEMEE CSCHLPYTEQ GMLAEDFKQ KALGLIQTAS RHAEVITPAV 1750 QTMWKLEVF WARHMWFTS GIQVLAGLST LPCNPAIASL MAFTAAVTSP 1800 LITTGOTILIFN ILGGWAAQL AAPCLATAFV GAGLAGAAIG SVGLGKVUD 1850	CCPERMASCR PIDWFAQGWG PITYIKPNSS DQRPYCWHYA PRPCGVVPAS	500
GCTWNSTGF TKTCGSPEN IGGVENRILI CPIDCFRKHP EATYTKOSS 600 PWLTPRCLVD YPYRLMYPC TLNFSIFKVR MYVGVEHRL NAACMWIRGE 650 RCNLEDRDRS ELSPILLSTT EWILLPCAFT TLPALSTGLI HIHONIVDVQ 700 YLYGVGSAFV SFAIKWEYIL LLFILLADAR VCACLIMMIL TAQAFAALDN 750 LWINAASVA GAHGILSFLV FFCAAWYIKG RLAFCAAYAF YGWPLILLL 800 LALPPRAYAL DTEVAASCGG VVLVGLMALT LSPYYKRYIS WCMWLLYFL 850 TRVEAQLHW VPPLWRGCR DAVILLMCVV HPTLVFDITK LLLAIFGPLW 900 HLQASLLKVP YFVRVQELLR ICALARKIAG GHYVMAIIK LGALTGIYVY 950 NHLTPLRDWA HNELRDLAVA VEPVVFSRME TKLITWADDI AACGDIINGL 1000 PVSARRGQEI LLGPADGW/S KGWRLLAPIT AYAQQTRGLL GCITTSLTGR 1050 DKNQVEGEVQ IVSTATQTFL ATCINGVCWI VYHAAGTRTI ASPKGPVIQM 1100 YTNVDQDLVG WPARQESRSL TPCTCGSSDL YLVTRHADVI PVRRRGDSRG 1150 SLLSPRPISY LKGSSGFLL CPACHAVCLF RAAVCTRGVA KAVDFIPVEN 1200 LGTTMRSPVF TINSSPPAVP QSFQVAHLHA PTGSCKSTKV PAAYAAQGYK 1250 VLVINPSVAA TLGFGAYMSK AHGVDPNIRT GVRTTTTGSP TTYSTYCKFL 1300 ADGGCSGGAY DILICDECHS TDATSILGIG TVLDQAETAG ARLWLATAT 1350 PPGSVIVSHP NIEEVALSTT GEIPFYGKAI PLEVIKGGRH LIFCHSKKKC 1400 DELAAKLVAL GINVAYYRG LDVSVIPTSG DVVVSTDAL MIGFTCDFDS 1450 VIDCNICVIQ TVFFSLDPIF TIETTTLPQD AVSRTQRRGR TGRCKFGIYR 1550 FVARGERESG MFDSSVLCEC YDAGCAWYEL TPAETTVRLR AYMNTPGLEV 1550 CQDHLEFWIG VFTGLTHIDA HFLSQIKQSG ENFFLVAYQ ATVCARAQAP 1600 PPSWDQMKC LIRLKPTLHG PTPLLYRIGA VQNEVILTHP ITKYTMICMS 1650 ADLEWTSTW VLVGGVLAAL AAYCLSTGCV VIVURIVLSG KPAITPDREV 1700 LYQEFDEMEE CSQHLFYTEQ CMM.AEDFKQ KALGLLQTAS RHAEVTTPAV 1750 GINWQKLEVF WAKHMINFIS GIQVLAGLST LPCNPAIASL MAFTAAVTSP 1800 LITTGOTILIFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD 1850		55 0
RCNIEDRDRS ELSPLISTI EWQILPCAFT TIPALSIGLI HIKANIVDAQ YLYGVGSAFV SFAIKWEYIL ILFILIADAR VCACLWMIL IAQAEAALEN 750 LWINAASVA GAHGILSFLV FFCAAWYIKG RLAPGAAYAF YGWFLILLL 800 LALPPRAYAL DTEVAASCOG VVIVGIMALT LSPYYKRYIS WCMWLQYFL 850 TRVEAQLHW VPPINVRGER DAVILIMCVV HPILVFDITK LLIAIFGPIW 900 HLQASLIKVP YFVRVQGLIR ICALARKIAG GHYVQMAIIK LGALIGIYVY 950 NHLTPLRWA HNGLRDLAVA VEPVVFSRME TKLTIVGADT AACCDIINGL 1000 PVSARRQQEI LLGPADGWS KGWRLIAPIT AYAQQTRGIL GCIITSLIGR 1050 DKNQVEGEVQ IVSTATQIFL ATCINGVCWI VYHGAGIRTI ASPKGPVIQM 1100 YTNVDQDLWG WPAPQGSRSL TPCICGSSDL YLVTRHADVI PVRRRCDSRG 1150 SLLSPRPISY LKGSSGGPLL CPACHAVGLF RAAVCIRGVA KAVDFIPVEN 1200 LGTIMRSFVF TINSSPPAVP QSFQVAHLHA PIGSGKSTKV PAAYAAQGYK 1250 VLVINPSVAA TLGFGAYMSK AHGVDRIRT GVRTTITGSP ITYSTYGKFL 1300 ADGCSGGAY DIIICDECHS TDATSILGIG TVLDQAETAG ARLWLATAT 1350 PPGSVIVSHP NIEEVALSTT GEIPFYGKAI PLEVIKOGRI LIFCHSKKC 1400 DELAAKLVAL GINAVAYRG LDVSVIPTSG DVVVSTDAL MIGSTGDFDS 1450 VIDCNICVIQ TVDFSLDPIF TIETITLPQD AVSRTQRRGR TGRGKFGTYR 1500 PPSWDQMKC LIRLKPTILG PTPLLYRIGA VQNEVILINE TTRKYIMICMS 1650 ADLEWISIW VLVGSVAAL AAYCLSTGCV VIVGRIVLSG KPALIPDREV 1700 LYQEFDEMEE CSQHLPYTEQ CMMLAEDFKQ KALGLLQTAS RHAEVITPAV 1750 CIMWCKLEVF WARMMNFIS GIQYLAGLST LPCNPAIASL MAFTAAVTSP 1800 LITTGOTILIFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVELGKVLVD 1850		600
YLYGVGSAFV SFAIKWEYIL LIFLIADAR VCACIMMIL TAQAFAALEN 750 LWINAASVA GAHGILSFLV FFCAAWYIKG RLAPGAAYAF YGWPLLILL 800 LALPPRAYAL DTEVAASCOG VVIVGIMALT LSPYYRYIS WCMWLQYFL 850 TRVEAQLHW VPPLWRGGR DAVILIMCVV HPILVFDITK LLLAIFGPLW 900 ILQASLLKVP YFVRVQGILR ICALARKIAG GHYVQMAIIK LGALIGIYVY 950 NHLTPLRUWA HNGLRDLAVA VEPVVFSRME TKLITWGADT AACCDIINGL 1000 PVSARRQQEI LLGPADGW/S KGWRLIAPIT AYAQQIRGIL GCIITSLIGR 1050 DKNQVEGEVQ IVSTATQIFL ATCINGVCMV VYHGAGIRTI ASPKGPVIQM 1100 YTNVDQDLVG WPAPQGSRSL TPCTCGSSDL YLVTRHADVI PVRRCDSRG 1150 SLLSPRPISY LKGSSGFLL CPACHAVGLF RAAVCTRGVA KAVDFIPVEN 1200 LGTIMRSPVF TINSSPPAVP QSFQVAHLHA PIGSGKSTKV PAAYAAQGYK 1250 VIVINPSVAA TLGFGAYMSK AHGVDENIRT GVRTITIGSP TTYSTYCKFL 1300 ADGGCSGGAY DIIICDECHS TDATSILGIG TVLDQAETAG ARLWLATAT 1350 PFGSVIVSHP NIEEVALSTT GEIPFYGKAI PLEVIKGGRH LIFCHSKKC 1400 DELAAKIVAL GINAVAYYRG LDVSVIPTSG DVVVVSTDAL MIGETGDFDS 1450 VIDCNICVIQ TVDFSLDPIF TIETTTILPQD AVSRTQRGR TGRCKPGIYR 1500 FVARGERPSG MFDSSVLCEC YDAGCAWYEL TPAETTVRLR AYMNTPGLFV 1550 CQDELEFWEG VFTGLIHIDA HFLSQTKQSG ENFPYLVAYQ ATVCARAQAP 1600 PPSNIDMMKC LIRLKPTLHG PTPLLYRLGA VQNEVTLIHP ITKYIMICMS 1650 ADLEVVISTW VLVGGVLAAL AAYCLSTGCV VTVGRIVLSG KPALIPDREV 1700 LYQEFDEMEE CSQHLPYIDQ GMMLAEQFKQ KALGILQTAS RHAEVITPAV 1750 QIIMQKLEVF WAKHMNIFIS GIQYLAGLST LPCNPAIASL MAFTAAVTSP 1800 LITTGOTLLFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVELEKVIVD 1850	PWLIPPCLVD YPYRLWHYPC TLNFSIFKVR MYVGGVEHRL NAACIWIRGE	650
YLYGVGSAFV SFAIKWEYIL ILIFILIADAR VCACLMMIL IAQAFAALEN 750 LWINAASVA GAHGILSFLV FFCAAWYIKG RLAPCAAYAF YGWPILIIL 800 LALPPRAYAL DIEVAASCOG VVIVGIMALT LSPYYKRYIS WMWILQYFL 850 TRVEAQLHW VPPLNVROGR DAVILIMCVV HPILVFDITK LLIAIFGPLW 900 ILQASILKVP YFVRVQEILR ICALARKIAG GHYVQMAIIK LGALIGIYVY 950 NHLTPLRUWA HNGLRDLAVA VEPVVFSRME TKLITWGADT AACCDIINGL 1000 PVSARROQEI LLGPADGMVS KGWRLIAPIT AYAQQIRGIL GCIITSLIGR 1050 DKNQVEGEVQ IVSTATQIFL ATCINGVOM VYHGAGIRTI ASPKGPVIQM 1100 YTNVDQDLVG WPAPQGSRSL TECTCGSSDL YLVTRHADVI PVRRCEDSRG 1150 SLLSPRPISY LKGSSGFPL CPACHAVGLF RAAVCIRGVA KAVDFIPVEN 1200 LGTIMRSPVF TINSSPPAVP QSFQVAHLHA PIGSGKSTKV PAAYAAQGYK 1250 VIVINPSVAA TICFGAYMSK AHGVDENIRT GVRTITIGSP TIYSTYCKFL 1300 ADGCCSGGAY DILICDECHS TDATSLIGIG TVLDQAETAG ARLVLATAT 1350 PFGSVIVSHP NIEEVALSTT GEIPFYGKAI PLEVIKGERH LIFCHSKKKC 1400 DELAAKLVAL GINAVAYYRG LDWSVIPTSG DVVWSTDAL MICFTEDFDS 1450 VIDCNICVIQ TVDFSLDPTF TIETTILPQD AVSRTQRGR TGRGKFGIYR 1500 FVAPGERPSG MFDSSVLCEC YDAGCAWYEL TPAETIVIR AYMNTPGLFV 1550 CQDHLEFWG VFTGLIHIDA HFLSQTKQSG BUFPYLVAYQ ATVCARAQAP 1600 PPSWDQMKC LIRLKPTLHG PTPLLYRIGA VQUEVILIHP TTKYIMICMS 1650 ADLEWTSTW VLVGGVLAAL AAYCLSTGCV VVCGRIVLSG KPALIPDREV 1700 LYQGFDEMEE CSQHLPYIDQ GMMLADQFKQ KALGILQTAS RHAEVTTPAV 1750 CTIMWOKLEVF WAKHMANFIS GIQYLAGLST LPCNPALASL MAFTAAVTSP 1800 LITTGOTLLFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVELEKVLVD 1850	RONLEDRORS ELSPLLISTT EWOILPCAFT TLPALSTGLI HLHQNIVDVQ	700
LWINAASVA GAKSILSFLV FFCAAWYIKG RIAPGAAYAF YGWPILLL 800 LALPPRAYAL DIEVAASCGG VVIWGIMALT LSPYYKRYIS WCMWLQYFL 850 TRVEAQLHW VPPINNRGER DAVILIMCVV HPILVFDITK LLLAIFGPLW 900 ILQASLLKVP YFVRVQGLIR ICALARKIAG GHYVQMAIIK LGALIGIYVY 950 NHLTPLRWA HNGIRDLAVA VEPVVFSRME TKLITWGADT AACGDIINGL 1000 PVSARRQQEI LLGPADGWVS KGWRLLAPIT AYAQQIRGLL GCIITSLIGER 1050 DKNQVEGEVQ IVSTATQIFL ATCINGVCWT VYHGAGIRTI ASPKGPVIQM 1100 YTNVDQDLVG WPAPQGSRSL TPCTCGSSDL YLVIRHADVI PVRRRCDSRG 1150 SLLSPRPISY LKGSSGGFLL CPACHAVGLF RAAVCTRGVA KAVDFIPVEN 1200 LGTIMRSPVF TINSSPPAVP QSFQVAHLHA PIGSGKSTKV PAAYAAQGYK 1250 VIMINPSVAA TLGFGAYMSK AHGVDENIRT GVRTTTIGSP ITYSTYGKFL 1300 ADCCCSGGAY DITICDECHS TDATSILGIG TVLDQAETAG ARLWLATAT 1350 PPGSVIVSHP NIEEVALSTT GEIPFYGKAI PLEVIKGGRH LIFCHSKKKC 1400 DELAAKIVAL GINAVAYYRG LDVSVIPTSG DVVVVSTDAL MIGFTGDFDS 1450 VIDCNICVIQ TVDFSLDPIF TIETTTLPQD AVSRTQRRGR TCRGKPGIYR 1500 FVARGERPSG MFDSSVLCEC YDAGCAWYEL TPAETIVRIR AYMNIPGLEV 1550 CQDHLEFWEG VFTGLIHIDA HFLSQIKQSG ENFPYLVAYQ ATVCARAQAP 1600 PPSWDQMKC LIRIKPTLHG PTPLLYRIGA VQUEVILTHP ITKYIMICMS 1650 ADLEVVTSIW VLVGGVLAAL AAYCLSTGCV VIVGRIVLSG KPALIPDREV 1700 LYQEFDEMEE CSQHLPYTEQ CMMLAEQFKQ KALGILQTAS RHAEVITPAV 1750 QINWQKLEVF WAKHMANFIS GIQYLAGLST LPCNPAIASL MAFTAAVTSP 1800 LTTGOTILFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVELGKVIVD 1850		750
TRVEAQLHW VPPINNRGER DAVILLMCW HPILVFDITK LLLAIFGPLW 900 ILQASLLKVP YFVRVQELR ICALARKIAG CHYVQMAIIK LGALIGIYVY 950 NHLTPLRWA HNEIRDLAVA VEPVVFSRME TKLITWGADT AACCDIINEL 1000 PVSARRQEI LLGPADGWS KGWRLLAPIT AYAQQIRGLL CCIITSLIGER 1050 DKNQVEGEVQ IVSTATQIFL ATCINGVCWT VYHGAGIRTI ASPKGPVIQM 1100 YTNVDQDLVG WPAPQGSRSL TPCTCGSSDL YLVIRHADVI PVRRRCDSRG 1150 SLLSPRPISY LKGSSGGPLL CPACHAVGLF RAAVCTRGVA KAVDFIPVEN 1200 LGTTMRSPVF TDNSSPPAVP QSFQVAHLHA PTGSGKSTKV PAAYAAQGYK 1250 VIMINPSVAA TLGFGAYMSK AHGVDRNIRT GVRTTTTGSP TTYSTYGKFL 1300 ADCCCSGGAY DITICDECHS TDATSILGIG TVLDQAETAG ARLWLATAT 1350 PPGSVIVSHP NIEEVALSTT GEIPFYGKAI PLEVIKGGRH LIFCHSKKKC 1400 DELAAKLVAL GINAVAYYRG LDVSVIPTSG DVVVVSTDAL MIGFTCDFDS 1450 VIDCNICVIQ TVDFSLDPTF TTETTTLPQD AVSRTQRRGR TCRGKRGIYR 1500 FVARGERPSG MFDSSVLCEC YDAGCAWYEL TPAETTVRIR AYMNIPGLEV 1550 CQDHLEFWEG VFTGLITHIDA HFLSQTKQSG ENFPYLVAYQ ATVCARAQAP 1600 PPSWDQMKC LIRLKPTLHG PTPLLYRIGA VQNEVILTHP TTKYIMICMS 1650 ADLEVVTSTW VLVGGVLAAL AAYCLSTGCV VIVGRIVLSG KPALIFDREV 1770 LYQEFDEMEE CSQHLPYTEQ CMMLAEQFKQ KALGILQTAS RHAEVITPAV 1750 QTINQKLEVF WAKHMNFIS GIQYLAGLST LPCNPAIASL MAFTAAVTSP 1800 LTTGOTILFN ILGGWAAQL AAPGAATAFV CAGLAGAAIG SVELGKVLVD 1850		800
TRVEAQLHW VPPLNRGGR DAVILLMCW HPTLVFDITK LLLAIFGPLW 900 ILQASLLKVP YFVRVQGLR ICALARKIAG GHYVQMAIK LGALIGIYVY 950 NHLTPLRWA HNGLRDLAVA VEPVVFSME TKLITWADT AACCDIINGL 1000 PVSARRQGEI LLGPADGMVS KGWRLLAPIT AYAQQTRGLL GCIITSLITGR 1050 DKNQVEGEVQ IVSTATQTFL ATCINGVCWT VYHCAGTRTI ASPKGPVIQM 1100 YTNVDQDLWG WPAPQGSRSL TPCTCGSSDL YLVTRHADUT PVRRRGDSRG 1150 SLLSPRPISY LKGSSGGPLL CPACHAVGLF RAAVCTRGVA KAVDFIPVEN 1200 LGTIMRSPVF TINSSPPAVP QSFQVAHLHA PTGSCKSTKV PAAYAAQGYK 1250 VIVINPSVAA TLGFGAYMSK AKGVDFNIRT GVRTTTTGSP TTYSTYCKFL 1300 ADGCCSGGAY DILICDECHS TDATSILGIG TVLDQAETAG ARLVVLATAT 1350 PPGSVIVSHP NIEEVALSTT GEIPFYGKAI PLEVIKGGRH LIFCHSKKKC 1400 DELAAKLVAL GINAVAYYRG LDVSVIPTSG DVVVVSTDAL MTGFTGDFDS 1450 VIDCNICVIQ TVDFSLDPTF TIETTTLPQD AVSRTQRRGR TCRCKPGIYR 1500 FVARGERPSG MFDSSVLCEC YDAGCAWYEL TPAETIVRIR AYMNTPGLPV 1550 CQDHLEFWEG VFTGLTHIDA HFLSQTKQSG ENFPYLVAYQ ATVCARAQAP 1600 PPSWDQMKC LIRLKPTLHG PTPLLYRLGA VQNEVILTHP TTKYIMTCMS 1650 ADLEVVISTW VLVGGVLAAL AAYCLSTGCV VIVGRIVLSG KPAITPDREV 1700 LYQEFDEMEE CSQHLPYTEQ CMMLAEDFKQ KALGLLQTAS RHAEVITPAV 1750 QINWQKLEVF WAKHMNFIS GIQYLAGLST LPCNPAIASL MAFTAAVTSP 1800 LITTGOTLLFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVELEKVLVD 1850		850
ILQASILKUP YFURUQELLR ICALARKIAG GHYUQMAIIK LGALTGIYUY 950 NHITPIRUMA HNGIRDIAVA VEPVUFSRME TKLITUGADT AACCDIINGL 1000 PVSARRQQEI LIGPADGMUS KGWRLIAPIT AYAQQIRGIL GCIITSLIGR 1050 DKNQVEGEVQ IVSTATQIFL ATCINGVCWT VYHGAGIRTI ASPKGFVIQM 1100 YTNVDQDLVG WPAPQGSRSL TPCTCGSSDL YLVIRHADVI FVRRGDSRG 1150 SILSPRPISY LKGSSGGPLL CPAGHAVGIF RAAVCTRGVA KAVDFIPVEN 1200 LGTIMRSPVF TINSSPPAVP QSFQVAHLHA PTGSGKSTKV PAAYAAQGYK 1250 VLVINPSVAA TLGFGAYMSK AHGVDFNIRT GVRTITTGSP HTYSTYGKFL 1300 ADGCCSGGAY DIIICDECHS TDATSILGIG TVLDQAETAG ARLWLATAT 1350 PPGSVTVSHP NIEEVALSTT GEIPFYGKAI PLEVIKGGRH LIFCHSKKC 1400 DELAAKLVAL GINAVAYYRG LDVSVIPTSG DVVVVSTDAL MIGFTGDFDS 1450 VIDCNICVIQ TVDFSLDPIF TIETTTILPQD AVSRTQRRGR TGRGKPGIYR 1500 FVAPGERPSG MFDSSVLCEC YDAGCAWYEL TPAETTVRLR AYMNTPGLPV 1550 CQDHLEFWEG VFTGLIHIDA HFLSQIKQSG ENFYLVAYQ ATVCARAQAP 1600 PPSWDQMKC LIRLKPTIHG PTPLLYRLGA VQNEVILIHP ITKYIMICMS 1650 ADLEVVTSIW VLVGGVLAAL AAYCLSTGCV VIVGRIVLSG KPALIPDREV 1700 LYQEFDEMEE CSQHLPYIEQ CMMLAEQFKQ KALGLLQTAS RHAEVITPAV 1750 QTIMQKLEVF WAKHMNIFIS GIQYLAGLST LPCNPALASL MAFTAAVTSP 1800 LTTGOTILFN ILGGWAAQL AARGAATAFV GAGLAGAAIG SVELGKVLVD 1850	TRVEAOLHW VPPLNVRGGR DAVILLMCVV HPTLVFDITK LLLAIFGPLW	900
NHITPLRIWA HIGLROLAVA VEPVVFSRME TKLTIWGADI AACCDIINGL PVSARRQQEI LLGPADGWS KGWRLLAPIT AYAQQIRGLL GCIITSLIGR DKNQVEGEVQ IVSTATQIFL ATCINGVCWI VYHGAGIRTI ASPKGPVIQM 1100 YTNVDQDLVG WPAPQGSRSL TPCTCGSSDL YLVTRHADVI PVRRGDSRG 1150 SILSPRPISY LKGSSGGPLL CPACHAVGLF RAAVCIRGVA KAVDFIPVEN 1200 LGTIMRSPVF TINSSPPAVP QSFQVAHLHA PIGSGKSIKV PAAYAAQGYK 1250 VLVINPSVAA TLGFGAYMSK AHGVDFNIRT GVRTITIGSP ITYSTYGKFL 1300 ADGCCSGGAY DILICIDECHS TDATSILGIG TVLDQAETAG ARLWLATAT 1350 PPGSVIVSHP NIEEVALSIT GEIPFYGKAI PLEVIKGGRH LIFCHSKKC 1400 DELAAKLVAL GINAVAYYRG LDVSVIPTSG DVVVVSIDAL MIGFTGDFDS VIDCNICVIQ TVDFSLDPIF TIETITLPQD AVSRTQRRGR TGRGKPGIYR 1500 FVAPGERPSG MFDSSVLCEC YDACCAWYEL TPAETIVRLR AYMNTRGLEV 1550 CQDHLEFWEG VFTGLIHIDA HFLSQIKQSG ENFPYLVAYQ ATVCARAQAP 1600 PPSWDQMWC LIRLKPITHG PTPLLYRLGA VQNEVILIHP ITKYIMICMS 1650 ADLEWISTW VLVGGVLAAL AAYCLSIGCV VIVGRIVLSG KPALIPDREV 1700 LYQEFDEMEE CSQHLPYIEQ GMMLAEQFKQ KALGILQTAS RHAEVITPAV 1750 QINWQKLEVF WAKHMNFIS GIQYLAGLST LPCNPALASL MAFTAAVTSP 1800 LTTGOTILIFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD	TIOASLIKUP YEVRVOGILR ICALARKIAG GHYVQMAIIK LGALIGIYVY	950
PVSARROQEI LLGPADEMVS KGWRLLAPIT AY-QQIRGLL GCIITSLIGR 1050 DKNQVEGEVQ IVSTATQIFL ATCINGVCWT VYHGAGIRTI ASPKGPVIQM 1100 YINVDQDLVG WPAPQGSRSL TPCTCGSSDL YLVIRHADVI PVRRGDSRG 1150 SILSPRPISY LKGSSGGPLL CPAGHAVGLF RAAVCIRGVA KAVDFIPVEN 1200 LGTIMRSPVF TDNSSPPAVP QSFQVAHLHA PIGSGKSTKV PAAYAAQGYK 1250 VLVINPSVAA TLGFGAYMSK AHGVDPNIRT GVRTITIGSP TIYSIYGKFL 1300 ADGCCSGGAY DIIICDECHS TDATSILGIG TVLDQAETAG ARLVVLATAT 1350 PRGSVIVSHP NIEEVALSTI GEIPFYGKAI PLEVIKGGRH LIFCHSKKKC 1400 DELAAKLVAL GINAVAYYRG LDVSVIPTSG DVVVVSTDAL MIGFTGDFDS 1450 VIDCNICVIQ TVDFSLDPIF TIETTTLFQD AVSRTQRRGR TGRGKPGIYR 1500 FVAPGERPSG MFDSSVLCEC YDAGCAWYEL TPAETIVRIR AYMNTRGLFV 1550 CQDHLEFWEG VFTGLIHIDA HFLSQIKQSG ENFPYLVAYQ ATVCARAQAP 1600 PPSWDQMWKC LIRLKPILHG PTPLLYRLGA VQNEVILHP ITKYIMICMS 1650 ADLEVVTSIW VLVGGVLAAL AAYCLSIGCV VIVGRIVLSG KPALIPDREV 1700 LYQEFDEMEE CSQHLPYIEQ GMMLABQFKQ KALGLIQTAS RHAEVITPAV 1750 QINWQKLEVF WAKHMNFIS GIQYLAGLST LPGNPATASL MAFTAAVTSP 1800 LTTGOTLLFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVELGKVLVD 1850	NHITPLROWA HNGLROLAVA VEPVVFSRME TKLTIWGADI AACGDIINGL	1000
DKNQVEGEVQ IVSTATQTFL ATCINGVCWT VYHGAGIRTI ASPKGPVIQM YTNVDQDLVG WPAPQGSRSL TPCTCGSSDL YLVTRHADVI PVRRGDSRG 1150 SILSPRPISY LKGSSGGPLL CPACHAVCLF RAAVCTRGVA KAVDFIPVEN 1200 LGTTMRSPVF TINSSPPAVP QSFQVAHLHA PTGSGKSTKV PAAYAAQGYK 1250 VLVINPSVAA TLGFGAYMSK AHGVDPNIRT GVRTTTTGSP ITYSTYGKFL 1300 ADGCCSGGAY DITICDECHS TDATSILGIG TVLDQAETAG ARLVVLATAT 1350 PPGSVTVSHP NIEEVALSTT GETPFYGKAI PLEVTKGGRH LTFCHSKKKC 1400 DELAAKLVAL GINAVAYYRG LDVSVIPTSG DVVVSTDAL MTGFTGDFDS VIDCNICVIQ TVDFSLDPTF TTETTTLPQD AVSRTQRRGR TGRGKPGTYR 1550 FVAPGERPSG MFDSSVLCEC YDAGCAWYEL TPAETTVRLR AYMNTPGLPV 1550 CQDHLEFWEG VFTGLTHIDA HFLSQTKQSG ENFPYLVAYQ ATVCARAQAP PPSWDQMKC LTRLKPTLHG PTPLLYRLGA VQNEVILTHP TTKYTMTCMS ADLEVVTSTW VLVGGVLAAL AAYCLSTGCV VTVGRTVLSG KPALTPDREV 1700 LYQEFDEMEE CSQHLPYTEQ CMMLAEQFKQ KALGLLQTAS RHAEVITPAV QTMQKLEVF WAKHMNFIS GTQYLAGLST LPCNPALASL MAFTAAVTSP 1800 LTTGOTLLFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD		1050
YINVDQDLVG WPAPQGSRSL TPCICGSSDL YLVIRHADVI PVRRRGDSRG 1150 SLLSPRPISY LKGSSGGPLL CPACHAVELF RAAVCIRGVA KAVDFIPVEN 1200 LGTIMRSPVF TIDNSSPPAVP QSFQVAHLHA PIGSGKSTKV PAAYAAQGYK 1250 VLVLNPSVAA TLGFGAYMSK AHGVDPNIRT GVRTITIGSP ITYSTYCKFL 1300 ADGCCSGGAY DILICIDECHS TDATSILGIG TVLDQAETAG ARLVVLATAT 1350 PPGSVIVSHP NIEEVALSTT GEIPFYGKAI PLEVIKGGRH LIFCHSKKKC 1400 DELAAKLVAL GINAVAYYRG LDVSVIPTSG DVVVVSTDAL MIGFTGDFDS 1450 VIDCNICVIQ TVDFSLDPIF TIETITLPQD AVSRTQRRGR TGRGKPGIYR 1500 FVARGERPSG MFDSSVLCEC YDAGCAWYEL TPAETITVRLR AYMNTPGLPV 1550 CQDHLEFWEG VFTGLTHIDA HFLSQIKQSG ENFPYLVAYQ ATVCARAQAP 1600 PPSWDQMKC LIRLKPTLHG PTPLLYRLGA VQNEVILTHP ITKYLMICMS 1650 ADLEVVTSIW VLVGGVLAAL AAYCLSTGCV VIVGRIVLSG KPALIPDREV 1700 LYQEFDEMEE CSQHLPYIEQ CMMLAEQFKQ KALGLLQTAS RHAEVITPAV 1750 QINWQKLEVF WAKHMNFIS GIQYLAGLST LPCNPAIASL MAFTAAVTSP 1800 LTTGOTLLFN ILCGWAAQL AAPGAATAFV GAGLAGAAIG SVELGKVLVD 1850	DKNOVEGEVO IVSTATOTFL ATCINGVOWT VYHGAGIRTI ASPKGPVIQM	1100
SILSPRPISY LKGSSGGPLL CPAGHAVGLE RAAVCIRGVA KAVDFIPVEN LGTIMRSPVF TINSSPPAVP QSFQVAHLHA PIGSGKSTKV PAAYAAQGYK VLVINPSVAA TLGFGAYMSK AHGVDPNIRT GVRTITIGSP ITYSIYGKFL ADGCCSGGAY DILICIDECHS TIATSILGIG TVLDQAETAG ARLVVLATAT 1350 PPGSVIVSHP NIEEVALSIT GEIPFYGKAI PLEVIKGGRH LIFCHSKKKC 1400 DELAAKLVAL GINAVAYYRG LDVSVIPTSG DVVVVSTDAL MIGFTGDFDS VIDCNICVIQ TVDFSLDPIF TIETITLPQD AVSRTQRRGR TGRGKPGIYR FVAPGERPSG MFDSSVLCEC YDACCAWYEL TPAETTVRLR AYMNTPGLPV 1550 CQDHLEFWEG VFTGLIHIDA HFLSQIKQSG ENFPYLVAYQ ATVCARAQAP PPSWDQMKC LIRLKPILHG PTPLLYRLGA VQNEVILTHP ITKYIMICMS ADLEVVISTW VLVGGVLAAL AAYCLSTGCV VIVGRIVLSG KPALIPDREV LYQEFDEMEE CSQHLPYIEQ GMMLAEQFKQ KALGILQTAS RHAEVITPAV 1750 QTIMQKLEVF WAKHMNFIS GIQYLAGLST LPCNPAIASL MAFTAAVTSP LTTGOTLLFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD 1850	YTNVDODLYG WPAPOGSRSL TPCTCGSSDL YLVTRHADVI PVRRGDSRG	1150
LGTIMRSPVF TINSSPPAVP QSFQVAHLHA PIGSGKSTKV PAAYAAQGYK VLVLNPSVAA TLGFGAYMSK AHGVDPNIRI GVRTITIGSP ITYSTYGKFL 1300 ADGGCSGGAY DILICDECHS TDATSILGIG TVLLQAETAG ARLVVLATAT 1350 PPGSVIVSHP NIEEVALSIT GEIPFYGKAI PLEVIKGGRH LIFCHSKKKC 1400 DELAAKLVAL GINAVAYYRG LDVSVIPTSG DVVVVSIDAL MIGFTGDFDS 1450 VIDCNICVIQ TVDFSLDPIF TIETITILPQD AVSRTQRRGR TGRGKPGIYR 1500 FVAPGERPSG MFDSSVLCEC YDAGCAWYEL TPAETIVRLR AYMNIPGLPV 1550 CQDHLEFWEG VFTGLIHIDA HFLSQIKQSG ENFPYLVAYQ ATVCARAQAP 1600 PPSWDQMWKC LIRLKPILHG PTPLLYRIGA VQNEVILITHP ITKYIMICMS 1650 ADLEVVTSIW VLVGGVLAAL AAYCLSTGCV VIVGRIVLSG KPAILPDREV 1700 LYQEFDEMEE CSQHLPYIEQ GMMLAEQFKQ KALGLLQTAS RHAEVITPAV 1750 QTIMQKLEVF WAKHMNIFIS GIQYLAGLST LPGNPAIASL MAFTAAVISP 1800 LTTGOTLLFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD	SLISPRPISY LKGSSGGPLL CPACHAVGLF RAAVCTRGVA KAVDFIPVEN	1200
VLVINPSVAA TLGFGAYMSK AHGVDENIRT GVRTITIGSP ITYSIYGKFL ADGOCSGGAY DIJICDECHS TDATSILGIG TVLDQAETAG ARLWLATAT 1350 PPGSVIVSHP NIEEVALSIT GEIPFYGKAI PLEVIKGGRH LIFCHSKKKC 1400 DELAAKLVAL GINAVAYYRG LDVSVIPTSG DVVVVSTDAL MIGFTGDFDS 1450 VIDCNICVIQ TVDFSLDPIF TIETTILEQD AVSRIQRRGR TGRGKFGIYR 1500 FVAPGERPSG MFDSSVLCEC YDAGCAWYEL TPAETIVRLR AYMNTPGLEV 1550 CQDHLEFWEG VFTGLIHIDA HFLSQTKQSG ENFPYLVAYQ ATVCARAQAP 1600 PPSWDQMKC LIRLKPTLHG PTPLLYRLGA VQNEVILTHP ITKYLMICMS 1650 ADLEVVISTW VLVGGVLAAL AAYCLSTGCV VIVGRIVLSG KPALIPDREV 1700 LYQEFDEMEE CSQHLPYTEQ GMMLAEQFKQ KALGLLQTAS RHAEVITPAV 1750 QTINQKLEVF WAKHMINFIS GIQYLAGLST LPGNPAIASL MAFTAAVTSP 1800 LTTGOTLLFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD		1250
ADGCCSGAY DILICIDECHS TDATSLIGIG TVLDQAETAG ARLVVLATAT 1350 PPGSVIVSHP NIEEVALSIT GEIPFYGKAI PLEVIKGGRH LIFCHSKKC 1400 DELAAKLVAL GINAVAYYRG LDVSVIPTSG DVVVVSIDAL MIGFTGDFDS 1450 VIDCNICVIQ TVDFSLDPIF TIETTILPQD AVSRTQRRGR TGRGKPGIYR 1500 FVAPGERPSG MFDSSVLCEC YDAGCAWYEL TPAETIVRLR AYMNTPGLFV 1550 CQDHLEFWEG VFTGLIHIDA HFLSQIKQSG ENFPYLVAYQ ATVCARAQAP 1600 PPSWDQMWKC LIRLKPTLHG PTPLLYRLGA VQNEVILIHP ITKYLMICMS 1650 ADLEVVISIW VLVGGVLAAL AAYCLSTGCV VIVGRIVLSG KPALIPDREV 1700 LYQEFDEMEE CSQHLPYIEQ GMMLAEQFKQ KALGLLQTAS RHAEVITPAV 1750 QINWQKLEVF WAKHMANFIS GIQYLAGLST LPGAPAIASL MAFTAAVTSP 1800 LTTGOTLLFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD 1850	VININPSVAA TIGFGAYMSK AHGVDPNIRT GVRTITIGSP ITYSTYGKFL	1300
PPGSVIVSHP NIEEVALSIT GEIPFYGKAI PLEVIKGGRH LIFCHSKKKC DELAAKLVAL GINAVAYYRG LDVSVIPISG DVVVVSTDAL MIGFTGDFDS 1450 VIDCNICVIQ TVDFSLDPIF TIEITILPQD AVSRTQRRGR TGRGKPGIYR FVAPGERPSG MFDSSVLCEC YDAGCAWYEL TPAETIVRLR AYMNIPGLPV 1550 CQDHLEFWEG VFTGLIHIDA HFLSQIKQSG ENFPYLVAYQ ATVCARAQAP 1600 PPSWDQMWKC LIRLKPILHG PIPLLYRLGA VQNEVILTHP ITKYIMICMS 1650 ADLEVVISTW VLVGGVLAAL AAYCLSTGCV VIVGRIVLSG KPALIPDREV LYQEFDEMEE CSQHLPYIEQ GMMLAEQFKQ KALGLLQTAS RHAEVITPAV 1750 QINWQKLEVF WAKHMNFIS GIQYLAGLST LPCNPALASL MAFTAAVTSP 1800 LTTGOTLLFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD		1350
DELAAKLVAL GINAVAYYRG LDVSVIPTSG DVVVVSTDAL MIGFTGDFDS 1450 VIDCNICVIQ TVDFSLDPIF TIETTILPQD AVSRTQRRGR TGRGKPGIYR 1500 FVAPGERPSG MFDSSVLCEC YDAGCAWYEL TPAETIVRLR AYMNTPGLPV 1550 CQDHLEFWBG VFTGLITHIDA HFLSQIKQSG ENFPYLVAYQ ATVCARAQAP 1600 PPSWDQMWKC LIRLKPILHG PTPLLYRLGA VQNEVILIHP ITKYIMICMS 1650 ADLEVVISTW VLVGGVLAAL AAYCLSTGCV VIVGRIVLSG KPALIPDREV 1700 LYQEFDEMEE CSQHLPYIEQ GMMLAEQFKQ KALGLLQTAS RHAEVITPAV 1750 QINWQKLEVF WAKHMINFIS GIQYLAGLST LPCNPALASL MAFTAAVTSP 1800 LTTGOTLLFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD 1850		1400
VIDONICVIQ TVDFSLDPIF TIEITILPQD AVSRTQRRGR TGRGKPGIYR 1500 FVAPGERPSG MFDSSVLCEC YDAGCAWYEL TPAETIVRLR AYMNIPGLPV 1550 CQDHLEFWEG VFTGLIHIDA HFLSQTKQSG ENFPYLVAYQ ATVCARAQAP 1600 PPSWDQMWKC LIRLKPILHG PIPLLYRLGA VQNEVILTHP ITKYIMICMS 1650 ADLEVVISTW VLVGGVLAAL AAYCLSTGCV VIVGRIVLSG KPAIIPDREV 1700 LYQEFDEMEE CSQHLPYIEQ CMMLAEQFKQ KALGLLQTAS RHAEVITPAV 1750 QTMWQKLEVF WAKHMNFIS GIQYLAGLST LPCNPAIASL MAFTAAVTSP 1800 LTTGOTLLFN ILCGWAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD 1850		1450
FVAPGERPSG MFDSSVLCEC YDAGCAWYEL TPAETIVRLR AYMNTPGLPV 1550 CQDHLEFWEG VFTGLIHIDA HFLSQIKQSG ENFPYLVAYQ ATVCARAQAP 1600 PPSWDQMWKC LIRLKPILHG PTPLLYRLGA VQNEVILIHP ITKYIMICMS 1650 ADLEVVISTW VLVGGVLAAL AAYCLSTGCV VIVGRIVLSG KPALIPDREV 1700 LYQEFDEMEE CSQHLPYIEQ CMMLAEQFKQ KALGLLQTAS RHAEVITPAV 1750 QINWQKLEVF WAKHMINFIS GIQYLAGLST LPCNPALASL MAFTAAVTSP 1800 LTTGOTLLFN ILCGWAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD 1850	VIDOVICVIO TVDFSLDPIF TIETITLPQD AVSRTQRRGR TGRGKPGIYR	1500
CODHLEFWEG VFTGLTHIDA HFLSQTKQSG ENFPYLVAYQ ATVCARAQAP 1600 PPSWDQMWKC LIRLKPTLHG PTPLLYRLGA VQNEVTLTHP TTKYIMTCMS 1650 ADLEVVTSTW VLVGGVLAAL AAYCLSTGCV VTVGRTVLSG KPAIIPDREV 1700 LYQEFDEMEE CSQHLPYTEQ CMMLAEQFKQ KALGLLQTAS RHAEVTTPAV 1750 QTMWQKLEVF WAKHMANFIS GIQYLAGLST LPCNPAIASL MAFTAAVTSP 1800 LTTGOTLLFN ILCGWAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD 1850	FVAPGERPSG MFDSSVLCEC YDAGCAWYEL TPAETIVRLR AYMNIPGLPV	1550
PPSWDQMWKC LIRLKPILHG PTPLLYRIGA VQNEVILHP ITKYIMICMS 1650 ADLEVVISIW VLVGGVLAAL AAYCLSIGCV VIVGRIVLSG KPAIIPDREV 1700 LYQEFDEMEE CSQHLPYIEQ GMMLAEQFKQ KALGLLQIAS RHAEVITPAV 1750 QINWQKLEVF WAKHMWNFIS GIQYLAGLSI LPCNPAIASL MAFTAAVISP 1800 LTIGOILLFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD 1850	CODHLEEWEG VETGLIHIDA HELSQIKQSG ENEPYLVAYQ ATVCARAQAP	1600
ADLEVVISIW VLVGGVLAAL AAYCLSIGCV VIVGRIVLSG KPAIIPDREV 1700 LYQEFDEMEE CSQHLPYIEQ CMMLAEQFKQ KALGLLQIAS RHAEVITPAV 1750 QINWQKLEVF WAKHMUNFIS GIQYLAGLSI LPCNPAIASL MAFTAAVISP 1800 LTTCOTLLFN ILCGWAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD 1850	PPSWDOMWKC LIRLKPILHG PIPLLYRLGA VQNEVILIHP ITKYIMIOMS	
LYQEFDEMEE CSQHLPYIEQ CMMLAEQFKQ KALGLLQTAS RHAEVITPAV 1750 QTMWQKLEVF WAKHMMNFIS GIQYLAGLST LPCNPAIASL MAFTAAVTSP 1800 LTTCOTLLFN ILCCWMAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD 1850	ADLEVVISTW VLVGGVLAAL AAYCLSTGCV VTVGRIVLSG KPAIIPDREV	1700
QINWOKLEVF WAKHMINFIS GIQYLAGLST LPCNPAIASL MAFTAAVTSP 1800 LTTCOTLLFN ILCCWAAQL AAPGAATAFV GAGLAGAAIG SVGLCKVLVD 1850	LYOFFDEMEE CSOHLPYIEQ CMMLAEQFKQ KALGLLQTAS RHAEVITPAV	
LTTGOTLLFN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD 1850	OINWOKLEVF WAKHMANFIS GIQYLAGLST LECENPAIASL MAFTAAVTSP	1800
ILAGYGAGVA GALVAFKUMS GEVPSTEDLV NILPATLSPG ALVVGVVCAA 1900	LTTCOTLLEN ILGGWAAQL AAPGAATAFV GAGLAGAAIG SVGLGKVLVD	
	ILAGYGAGVA GALVAFKINS GEVPSTEDLV NILPAILSPG ALVVGVVCAA	1900

FIG. 16G

SUBSTITUTE SHEET (RULE 26)

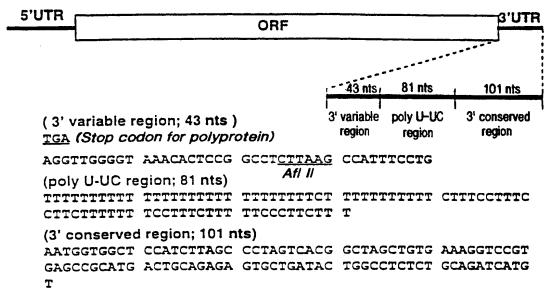
H77CV-J4aa Sequence

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TT PPHY/CPCE	CAVOWMNRLI	AFASRGNHVS	PIHYVPESDA	AARVIAILSS	1950
TITATOLLERI.	HOWISSECTT	PCSGSWLRDI	WDWICEVLSD	FKIWLKAKLM	2000
POT PETPET/S	CORGYRGWIR	GDGIMHIRCH	CGAELIGHVK	NGIMRIVGPR	2050
TONIMACTIF	PTNAYTIGEC	TPLPAPNYKF	ALWRVSAEEY	VEIRRVGDFH	210 0
WINDSMITTING.	KCPCOTPSPE	FFTELDGVRL	HRFAPPCKPL	LREEVSFRVG	2150
THEVENICON.	PCEPEPDVAV	LISMLIDPSH	ITAFAACRRL	ARGSPPSMAS	2200
CCACOT CAPS	TIKATYTANHD	SPDAELIEAN	LLWROEMGGN	TIRVESENKV	2250
2242 Denied IV	A E E DEREVSV	PAETLRKSRR	FARALPWAR	PDYNPPLVET	2300
	VHGCPI.PPPR	SPPVPPPRKK	RIVVLIESIL	STALAELATK	2350
CECCCOTOCT	TENTITISSE	PAPSGCPPDS	DVESYSSMPP	LEGEPGDPDL	2400
CTC CAICTLICS	CADTEDAVOC	SMSYSWIGAL	VIPCAAEEQK	LPINALSNSL	2450
	TSBSACOROK	KVIFDRLOVL	DSHYQDVLKE	VKAAASKVKA	2500
TEATTON TO T	TODOHSVKZK	FGYGAKDVRC	HARKAVAHIN	SWKDLLEDS	2550
VILLSVEEPLS	WIET TO TOPE	KGGRKPARLI	VFPDLGVRVC	EKMALYDWS	2600
ATAIDITIIM ATAIDITIIM	ACEUACECUE VIVENICAĞII	VEELVOAWKS	KKTPMEFSYL	TRCFDSIVIE	2650
KLPLAVIGS:	S TOTATIONS	VATKSITERI	YVGGPLINSF	GENOGYRROR	2700
SDIKIEEAL		A ACRAAGIOI	CIMLVOGDDI	WICESAGVQ	2750
ASGVLITSCO	a mandacabi	COPPOPEYDI	, FLITSCSSN	SVAHDGAGKR	2800 ~
EDAASLRAF	L EMAILLESSE	DELLOTED TO LOT	NITMFAPIL	ARMILMIHEF	2850
AAARIKDAI.	I. ETWKWWARET	A COSTEDIUM.	PITORLHGLS	AFSLHSYSPG	2900
SVLIARDQL	E CALINCELICA	A DUDARCIRAI	RIJSROGRAA	I CGKYLFNWAV	2950
EINRVAACL	K KIGALLIKU	N THINTOGEN	T VHSVSHARPI	R WFWFCILLLA	3000
		W FIRMINGE			3011
AGVGIYLLP	NR				

FIG. 16H

#1a. 3' Deletion mutants of pCV-H77C

Sequence of 3' untranslated region of pCV-H77C



#1a -1. pCV-H77C(-98X); 3' 98 nucleotides removed from pCV-H77C

TGAAGGTTGG GGTAAACACT CCGGCCTCTT AAGCCATTTC CTGTTTTTTT
TTTTTTTTTT TTTTTTTTTT TCTTTTTTT TTTCCTTTCCTTT
TTTTCCTTC TTTTTCCCTT CTTTAAT

#1a -2. pCV-H77C(-42X); 3' 42 nucleotides removed from pCV-H77C

#1a -3. pCV-H77C(X-52); All of the 3' UTR sequence, except 3' 49 nucleotides, removed from pCV-H77C

TGAGCCGCAT GACTGCAGAG AGTGCTGATA CTGGCCTCTC TGCAGATCAT

FIG. 17A

#1a -4. pCV-H77C(X); All of the 3' UTR sequence, except 3' 101 nucleotides, removed from pCV-H77C

TGARATGGTG GCTCCATCTT AGCCCTAGTC ACGGCTAGCT GTGARAGGTC
CGTGAGCCGC ATGACTGCAG AGAGTGCTGA TACTGGCCTC TCTGCAGATC
ATGT

#1a -5. pCV-H77C(+49X); The proximal 49 nucleotides of the 3' conserved region (98 nucleotides; AAT not included) removed from pCV-H77C

#1a -6. pCV-H77C(VR-24); First 24 nucleotides of the 3' variable region removed from pCV-H77C

#1a -7. pCV-H77C(-U/UC); Poly U-UC region removed from pCV-H77C

TGAAGGTTGG GGTAAACACT CCGGCCTCTT AAGCCATTTC CTGAATGGTG

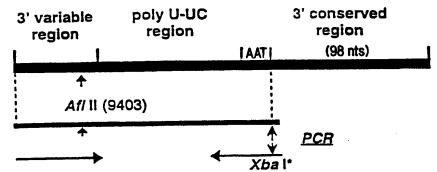
GCTCCATCTT AGCCCTAGTC ACGGCTAGCT GTGAAAGGTC CGTGAGCCGC

ATGACTGCAG AGAGTGCTGA TACTGGCCTC TCTGCAGATC ATGT

FIG. 17B

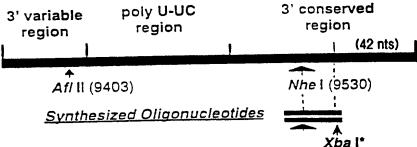
#1b. Strategy of 3' Deletion mutants

#1b -1. pCV-H77C(-98X)



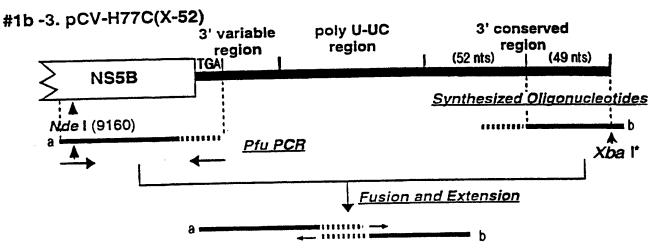
- 1. PCR Amplification
- 2. Purification of PCR products
- 3. Digestion with Afl II and Xba I
- 4. Cloning of Afl II IXba I fragment into pCV-H77C
- 5. Complete sequence analysis
- 6. in vitro transcription (within 24 hours of inoculation)
- 7. Percutaneous intrahepatic transfection into chimpanzee; 11/26/97 and 12/17/97
- 8. Result: Negative (No replication)

#1b -2. pCV-H77C(-42X)



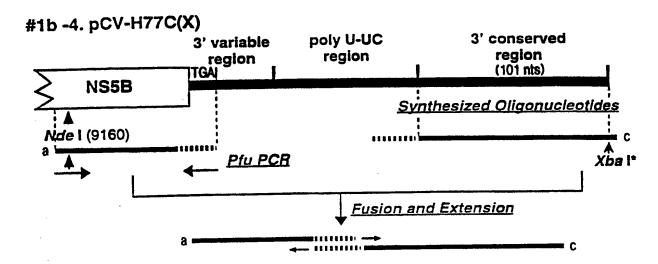
- 1. Synthesis of oligonucleotides (sense and anti-sense)
- 2. Hybridization of oligonucleotides
- 3. Digestion with Nhe I and Xba I
- 4. Cloning of Nhe I /Xba I fragment into pG9-KL26 (3' UTR of H77C)
- 5. Sequence analysis
- 6. Cloning of 3' UTR (-42X) [Afl II IXba I fragment] into pCV-H77C
- 7. Complete sequence analysis
- 8. in vitro transcription (within 24 hours of inoculation)
- 9. Percutaneous intrahepatic transfection into chimpanzee (Schedule; 1/22/98, 2/5/98)

FIG. 17C



- 1. Fragment a; Pfu PCR amplification and purification
- 2. Fragment b; Synthesized oligonucleotides (anti-sense)
- 3. Fusion and extension
- 4. TA cloning
- 5. Sequence analysis
- 6. Cloning Nde I-Xba I fragment with correct sequence into pCV-H77C
- 7. Complete sequence analysis
- 8. In vitro transcription (within 24 hours of inoculation)
- 9. Percutaneous intrahepatic transfection into chimpanzee

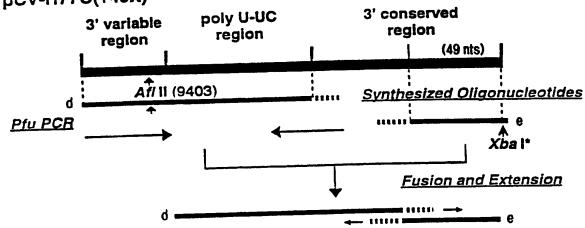
FIG. 17D



- 1. Fragment a; Pfu PCR amplification and purification
- 2. Fragment c; Synthesized oligonucleotides (anti-sense)
- 3. Fusion and extension
- 4. TA cloning
- 5. Sequence analysis
- 6. Cloning Nde I-Xba I fragment with correct sequence into pCV-H77C
- 7. Complete sequence analysis
- 8. In vitro transcription (within 24 hours of inoculation)
- 9. Percutaneous intrahepatic transfection into chimpanzee

FIG. 17E

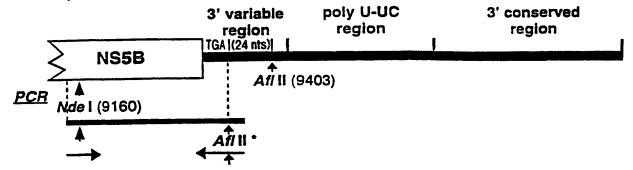
#1b -5. pCV-H77C(+49X)



- 1. Fragment d; Pfu PCR amplification and purification
- 2. Fragment e ; Synthesized oligonucleotides (anti-sense)
- 3. Fusion and extension
- 4. TA cloning
- 5. Sequence analysis
- 6. Cloning Afl II-Xba I fragment with correct sequence into pCV-H77C
- 7. Complete sequence analysis
- 8. In vitro transcription (within 24 hours of inoculation)
- 9. Percutaneous intrahepatic transfection into chimpanzee

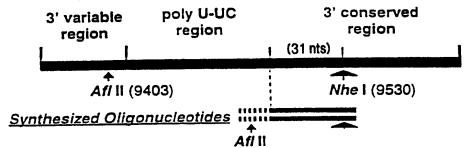
FIG. 17F

#1b -6. pCV-H77C(VR-24)



- 1. PCR Amplification
- 2. Purification of PCR products
- 3. Digestion with Nde I and Afl I
- 4. Cloning of Nde I /Afl II fragment into pCV-H77C
- 5. Complete sequence analysis
- 6. in vitro transcription (within 24 hours of inoculation)
- 7. Percutaneous intrahepatic transfection into chimpanzee

#1b -7. pCV-H77C(-U/UC)



- 1. Synthesis of oligonucleotides (sense and anti-sense)
- 2. Hybridization of oligonucleotides
- 3. Digestion with Aff II and Nhe I
- 4. Cloning of Afl II and Nhe I fragment into pG9-KL26
- 5. Sequence analysis
- 6. Cloning of 3' UTR (-poly U-UC) [Afl II /Xba I fragment] into pCV-H77C
- 7. Complete sequence analysis
- 8. in vitro transcription (within 24 hours of inoculation)
- 9. Percutaneous intrahepatic transfection into chimpanzee

FIG. 17G

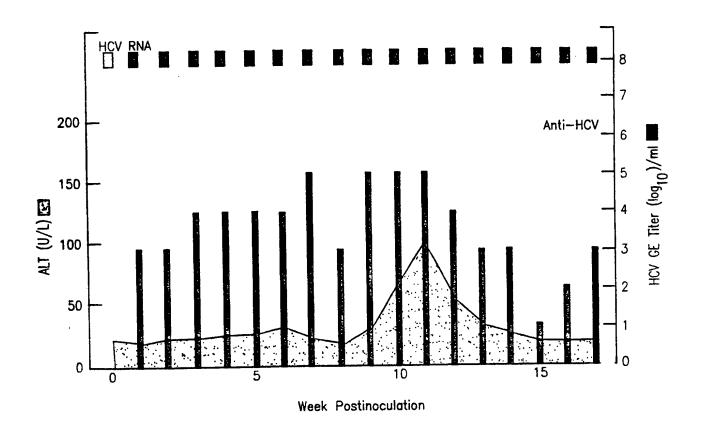


FIG. 18A

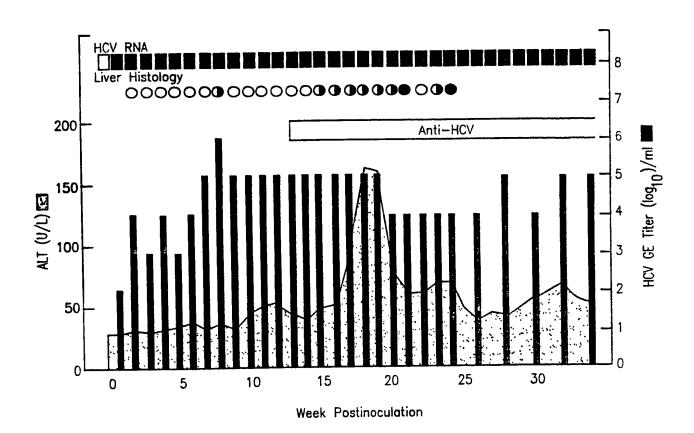


FIG. 18B

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) Internati nal Patent Classificati n 6: C12N 15/40, C07K 14/18, C12Q 1/70.

Α3

(11) International Publication Number:

WO 99/04008

C07K 16/10, A61K 39/29

(43) International Publication Date:

28 January 1999 (28.01.99)

(21) International Application Number:

PCT/US98/14688

(22) International Filing Date:

16 July 1998 (16.07.98)

(30) Priority Data:

60/053,062 09/014,416

18 July 1997 (18.07.97) US 27 January 1998 (27.01.98)

US

(71) Applicant: THE GOVERNMENT OF THE UNITED STATES OF AMERICA, asrepresented by THE SECRETARY, DEPARTMENT OF HEALTHAND HUMAN SERVICES [US/US]; Office of Technology Transfer, National Institutes | Published of Health, Suite 325, 6011 Executive Boulevard, Rockville, MD 20852 (US).

(72) Inventors: YANAGI, Masayuki; 257 Congressional Lane, #402, Rockville, MD 20852 (US). BUKH, Jens; 7 Center Drive, MSC 0740, Bethesda, MD 20892 (US). EMERSON, Suzanne, U.; 18201 Woodcrest Drive, Rockville, MD 20852 (US). PURCELL, Robert, H.; 17517 White Grounds Road, Boyds, MD 20841 (US).

(74) Agents: FEILER, William, S. et al.; Morgan & Finnegan, L.L.P., 345 Park Avenue, New York, NY 10154 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

With international search report.

(88) Date of publication of the international search report: 1 April 1999 (01.04.99)

(54) Title: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND USES THEREOF

(57) Abstract

The present invention discloses nucleic acid sequences which encode infectious hepatitis C viruses and the use of these sequences, and polypeptides encoded by all or part of these sequences, in the development of vaccines and diagnostics for HCV and in the development of screening assays for the identification of antiviral agents for HCV.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of Americ
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Internatio \pplication No PCT/US 98/14688

CLASSIFICATION OF SUBJECT MATTER C 6 C12N15/40 C07k IPC 6 A61K39/29 C07K16/10 C07K14/18 C12Q1/70According to International Patent Classification (IPC) or to both national classification and IPC **B. FIELDS SEARCHED** Minimum documentation searched (classification system followed by classification symbols) C12N C07K C12Q A61K IPC 6 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practical, search terms used) C. DOCUMENTS CONSIDERED TO BE RELEVANT Relevant to claim No. Citation of document, with indication, where appropriate, of the relevant passages Category 5 YOO B J ET AL: "Transfection of a 1,13-24. X 33 - 35. differentiated human hepatoma cell line 42.43 (Huh7) with in vitro-transcribed hepatitis C virus (HCV) RNA and establishment of a long-term culture persistently infected with HCV" JOURNAL OF VIROLOGY, vol. 69, no. 1, January 1995, pages 32-38, XP002022696 AMERICAN SOCIETY FOR MICROBIOLOGY US see the whole document Patent family members are listed in annex. X Further documents are listed in the continuation of box C. Special categories of cited documents: "T" later document published after the international filing date or priority date and not in conflict with the application but "A" document defining the general state of the art which is not considered to be of particular relevance cited to understand the principle or theory underlying the invention earlier document but published on or after the international "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "O" document referring to an oral disclosure, use, exhibition or document published prior to the international filing date but "&" document member of the same patent family later than the priority date claimed Date of mailing of the international search report Date of the actual completion of the international search 12 January 1999 22/01/1999 Authorized officer Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016 Cupido, M

1

Internatic Application No
PCT/US 98/14688

		PC1/US 98/14688				
C.(Continu	C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT					
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.				
E	WO 98 39031 A (UNIVERSITY OF WASHINGTON; KOLYKHALOV A; RICE C) 11 September 1998 see SEQ ID NO:1, representing a HCV H77 consensus sequence having 99.6% identity with the nucleic acid sequences in figures 4A-4F of the present application.	1,4,5, 13-24, 33-43				
X	EP 0 516 270 A (JAPAN IMMUNO INC) 2 December 1992 see the whole document	23,24, 40,41				
Α	WO 97 08310 A (WASHINGTON UNIVERSITY) 6 March 1997 see page 15; figure 3	1-43				
P,X	YANAGI M ET AL: "Transcripts of a chimeric cDNA clone of hepatitis C virus genotype 1b are infectious in vivo" VIROLOGY, vol. 244, no. 1, 25 April 1998, pages 161-172, XPO02089701 ORLANDO US see the whole document	1,6-10, 13-24, 33-43				

Intern unal application No.

PCT/US 98/14688

Box I Observations wher certain claims were found unsearchable (Continuation of item 1 of first sheet)
This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:
1. X Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely: Remark: Although claim(s) 41 and 43 is(are) directed to a method of treatment of the animal body, the search has been carried out and based on the alleged effects of the composition. 2. X Claims Nos.: 29 and 30
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically: The scope of claims 29 and 30 is so unclear and not well specified that a meaningful search was not possible
3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
B x II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)
This International Searching Authority found multiple inventions in this international application, as follows:
As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Remark on Protest The additional search fees wer accompanied by the applicant's protest. No protest accompanied the payment of additional search fees.

information on patent family members

Internatio	pplication No
PCT/US	98/14688

Patent document Publication cited in search report date		Publication date	Patent family member(s)		Publication date
WO 9839031	Α	11-09-1998	AU	6938698 A	22-09-1998
EP 0516270	Α	02-12-1992	JP CA	5091884 A 2065690 A	16-04-1993 11-10-1992
WO 9708310	Α	06-03-1997	AU CA EP	6909796 A 2230452 A 0856051 A	19-03-1997 06-03-1997 05-08-1998